Machine Learning Based Energy Vector Forecasting for Energy Management

Dissertation submitted to
Visvesvaraya National Institute of Technology, Nagpur
in partial fulfilment of requirement for the award of
the degree

Master of Technology
in
Integrated Power System

By

Nidhi Haribhau Gaikwad
(MT23IPS004)

Under the guidance of

Dr. Arghya Mitra

Department of Electrical Engineering
Visvesvaraya National Institute of Technology
Nagpur-440010(India)

(June2025)

Machine Learning Based Energy Vector Forecasting for Energy Management

Dissertation submitted to
Visvesvaraya National Institute of Technology, Nagpur
in partial fulfilment of requirement for the award of
the degree

Master of Technology
in
Integrated Power System

By

Nidhi Haribhau Gaikwad
(MT23IPS004)

Under the guidance of

Dr. Arghya Mitra

Department of Electrical Engineering
Visvesvaraya National Institute of Technology
Nagpur-440010(India)

(June 2025)

© Visvesvaraya National Institute of Technology (VNIT) 2025

Department of Electrical Engineering

Visvesvaraya National Institute of Technology, Nagpur

Declaration

I, **Nidhi Haribhau Gaikwad** hereby declare that this dissertation titled **"Machine Learning Based Energy Vector Forecasting for Energy Management"** is carried out by me in the Department of Electrical Engineering of Visvesvaraya National Institute of Technology, Nagpur. The work is original and has not been submitted earlier whole or in part for the award of any degree/diploma at this or any other Institution/University.

Gaikwad

Nidhi Haribhau Gaikwad

Enrolment No. MT23IPS004

Date:

Certificate

This is to certify that the dissertation titled **"Machine Learning Based Energy Vector Forecasting for Energy Management"** is submitted by **Nidhi Haribhau Gaikwad**, in partial fulfilment of the requirements for the award of the degree of **Master of Technology in Integrated Power System, VNIT Nagpur**. The work is comprehensive, complete, and fit for final evaluation.

Dr. Arghya Mitra

Supervisor, Assistant Professor

Department of Electrical Engineering

VNIT, Nagpur

Dr. B.S.Umre

Head, Department of Electrical Engineering,

VNIT, Nagpur

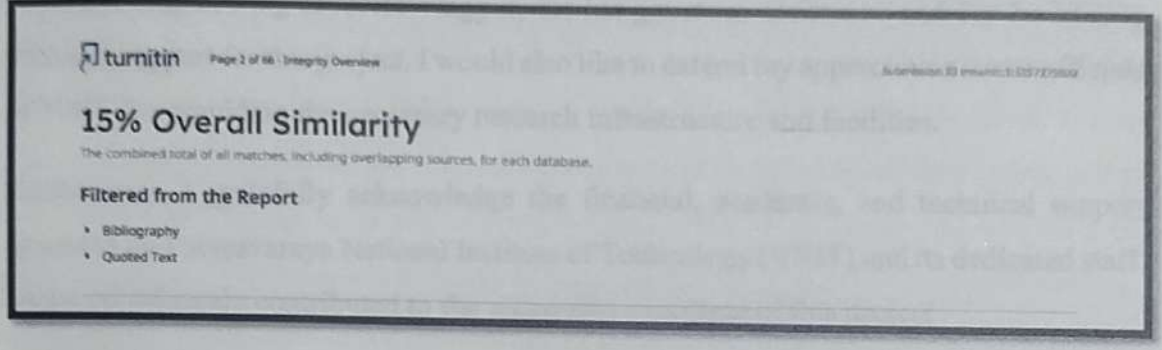
Date: 30.05.2025

Department of Electrical Engineering

Visvesvaraya National Institute of Technology, Nagpur

NON-PLAGIARISM CERTIFICATE

This is to certify that the M. Tech. Dissertation titled **"Machine Learning Based Energy Vector Forecasting for Energy Management"** submitted by **Nidhi Haribhau Gaikwad, MT23IPS004, ID No. 30794** is an original work. The similarity index is identified as **15%** which is within prescribed limits.


turnitin
Page 2 of 66 · Integrity Overview
15% Overall Similarity
The combined total of all matches, including overlapping sources, for each database.
Filtered from the Report
Bibliography
Quoted Text

Gaikwad

Nidhi Haribhau Gaikwad
MT23IPS004
MTech Integrated Power System
Department of Electrical Engineering
VNIT, Nagpur
Date

Dr. Arghya Mitra
Supervisor, Assistant Professor
Department of Electrical Engineering
VNIT, Nagpur

ACKNOWLEDGEMENTS

I would like to express my sincere and heartfelt gratitude to my supervisor, **Dr. Arghya Mitra**, for his invaluable guidance, constant encouragement, timely suggestions, and constructive feedback throughout the course of my research. His willingness to clarify my doubts and his continuous support played a crucial role in the successful completion of this work.

I am deeply indebted to **Dr. B. S. Umre**, Professor and Head of the Department of Electrical Engineering at VNIT Nagpur, for his generous assistance and for facilitating financial support for this project. I would also like to extend my appreciation to the officials of VNIT for providing the necessary research infrastructure and facilities.

Furthermore, I gratefully acknowledge the financial, academic, and technical support extended by Visvesvaraya National Institute of Technology (VNIT) and its dedicated staff, which significantly contributed to the successful execution of this project

I would like to express my sincere gratitude to Mr. Abhijeet Geetey and Mr. Vrundesh Pawde for their unwavering friendship and support. Their valuable advice and insightful suggestions served as a great source of motivation throughout my journey. I am also deeply thankful to research scholar Mr. Bimal Kumar Dora and Mrs. Payal Jangilwar for their continuous technical support and guidance during the course of my study

Finally, I owe my gratitude to my family members, friends, and classmates for their occasional support, suggestions, and help in making this project a success.

Nidhi Haribhau Gaikwad

ABSTRACT

With global energy consumption rising annually by over 4%, the transition to clean and sustainable energy has become a critical priority. Conventional non-renewable energy sources such as coal, oil, and natural gas are major contributors to environmental degradation and global warming due to excessive greenhouse gas emissions. Renewable energy sources (RES), particularly wind and solar, offer a promising solution, but their intermittent and unpredictable nature presents significant challenges for stable integration into power systems. Accurate forecasting of renewable energy generation and electricity demand is essential to ensure grid stability and efficient energy management, especially in grid-isolated regions. This study presents a comprehensive approach using Machine Learning (ML) and Deep Learning (DL) techniques for wind speed, solar irradiance, and load forecasting, followed by energy management using a Battery Energy Storage System (BESS). The research is carried out in three phases. In the first phase, four ML models—Multiple Linear Regression (MLR), Support Vector Regression (SVR), Random Forest Regression (RFR), and Decision Trees (DT) are applied for wind speed forecasting using data from the National Solar Radiation Database (NSRD) spanning January 2018 to December 2019. The models are evaluated based on RMSE, MSE, MAE, and R^2 metrics. Results indicate that data preprocessing significantly enhances model performance. SVR proves to be the most accurate, especially under seasonal variations, outperforming other models across all metrics. The second phase involves forecasting wind speed and solar irradiance using DL models XG Boost, Convolutional Neural Networks (CNN), and Long Short-Term Memory (LSTM) networks applied to data obtained from the Sol cast API for Goa's southern region. Among the models, LSTM shows superior performance, achieving an RMSE of 0.0984 and R^2 nearing 0.99, confirming its robustness in handling time-series data. In the final phase, load demand forecasting is performed using the same DL models, with LSTM again emerging as the most accurate. The peak load is estimated at 1,180 kW, and a hybrid system comprising an 800-kW wind turbine and a 500-kW solar PV array is proposed. To address power deficits from renewable intermittency, a 125 kW BESS is integrated and evaluated under three initial states of charge (SOC), effectively enhancing system stability.

LIST OF FIGURES

Figure 3.1. Flowchart to follow the method 22

Figure 3.2 Wind speed over the time period 23

Figure 3.3 Correlation Matrix 24

Figure 3.4.Scatter graph for the wind speed forecasting for case I 26

Figure 3.5. Scatter graph for the wind speed forecasting for case II 27

Figure 3.6 Flowchart for processing the proposed models 30

Figure 3.7 Scatter graph for the wind speed forecasting for (a) XG BOOST (b) CNN (c) LSTM 32

Figure 4.1 Flowchart for processing of proposed models 35

Figure 4.2 Scatter graph for Solar Irradiance Forecasting for (a) XGBOOST (b) CNN (c) LSTM 37

Figure 5.1 Flowchart of proposed methodology 43

Figure 5.2 Scatter graph for Load Demand Forecasting for (a) XGBOOST (b) CNN (c) LSTM 45

Figure 5.3 Represents the hybrid renewable energy system integrating solar wind and battery storage 47

Figure 5.4 Graphical representaion of (a) solar and wind generation with load demand profile foe 24 hrs (b) total generation of load demand for 24 hr period (c) deficit power 49

LIST OF TABLES

Table 3.1 Without Data Preprocessing ……. 25

Table 3.2 With Data Preprocessing ……. 27

Table 3.3 Seasonal Analysis with Data Preprocessing……. 28

Table 3.4 Error Metrics of Wind Speed Forecasting ……..……31

Table 4.1 Error Metrics of Solar Irradiance Forecasting …….……..36

Table 5.1 Performance Indices for Load Demand Forecasting …….44

NOMINCLATURE

RES – Renewable Energy Sources

ML – Machine Learning

DL – Deep Learning

WECS – Wind Energy Conversion System

EMS – Energy Management System

BESS – Battery Energy Storage

ELF – Electric Load Forecasting

PV – Photovoltaic

LOF – Local Outlier Factor

MLR – Multiple Linear Regression

SVR – Support Vector Regression

RFR – Random Forest Regression

DT – Decision Tress

NSRDB – National Solar Radiation Database

DHI -Diffuse Horizontal Irradiance

GHI – Global Horizontal Irradiance

RMSE – Root Mean Square Error

MAE – Mean Absolute Error

MSE – Mean Square Error

CNN – Convolutional Neural Network

XG BOOST – X Gradient Boost

LSTM – Long Short-Term Memory

SOC – State of Charge

PUBLICATIONS

1. Nidhi Haribhau Gaikwad, Payal Jangilwar, Soumyabrata Das, Arghya Mitra, "Impact of Data Preprocessing on Wind Forecasting Using Machine Learning Techniques" Published in 3rd IEEE International Conference on Power, Control and Computing Technologies **(ICPC²T 2024),** Raipur, India
2. Nidhi Haribhau Gaikwad, Payal Jangilwar, Bimal Kumar Dora, Soumyabrata Das, Arghya Mitra, "Deep Learning-Based Solar and Wind Energy Forecasting for Grid-Independent Systems" Accepted for publication in 5th IEEE International Conference on Sustainable Energy and Future Electric Transportation **(SEFET2025)**, Jaipur, India, Jul 2025.
3. Nidhi Haribhau Gaikwad, Bimal Kumar Dora, Payal Jangilwar, Mriduwani Verma Soumyabrata Das, Arghya Mitra, Sunil Bhatt, Ashwin Dabhale, "Machine Learning-Based Load Forecasting and Battery Energy Management for Efficient Renewable Integration", Accepted for publication in 5th IEEE International Conference on Sustainable Energy and Future Electric Transportation **(SEFET2025)**, Jaipur, India, Jul 2025.
4. Bimal Kumar Dora, Nidhi Haribhau Gaikwad, Pranjali Kulkarni, Tarun Gangadhar, Ruchita Waghmare, Sudip Halder, Arghya Mitra, Damien Ernst, Sunil Bhat, "Cost Optimization of Renewable Energy Installation in a Four-Microgrid System Using GBOML", Accepted for publication in 5th IEEE International Conference on Sustainable Energy and Future Electric Transportation **(SEFET2025)**, Jaipur, India, Jul 2025.
5. Kumari Kavya, Bimal Kumar Dora, Nidhi Haribhau Gaikwad, Shantanu Choudhary, Ruchita Waghmare, Sudip Halder, Ishan Srivastava, Arghya Mitra, Damien Ernst, Sunil Bhat, "Optimizing Renewable Energy Integration in Microgrids Using Graph-Based Optimization Modeling Language", Accepted for publication in 5th IEEE International Conference on Sustainable Energy and Future Electric Transportation **(SEFET2025)**, Jaipur, India, Jul 2025.

Contents

ABSTRACT ..... **i**

LIST OF FIGURES ..... **ii**

LIST OF TABLES ..... **iii**

NOMINCLATURE ..... **iv**

PUBLICATIONS ..... **v**

CHAPTER 1 ..... **2**

INTRODUCTION **2**

1.1 Background **2**

1.1.1 Need for Solar and Wind Forecasting with Load Prediction: 6

1.2 Problem Statement **7**

1.3 Objectives of the Study **7**

1.4 Scope of the Study **8**

1.5 Significance of the Study **8**

1.6 Thesis Organization **8**

CHAPTER 2 ..... **11**

BRIEF DESCRIPTION OF VARIOUS MACHINE LEARNING TOOLS **11**

2.1 Overview of Various machine learning tools **11**

2.1.1 Multiple Linear Regression (MLR) 11

2.1.2 Random Forest Regression (RFR) 12

2.1.3 Decision Tree (DT) 12

2.1.4 Support Vector Regression (SVR) 13

2.1.5 Convolutional Neural Network (CNN) 14

2.1.6 XGBOOST 15

2.1.7 Long Short-Term Memory (LSTM) 16

2.2 Implementation of Machine Learning Model **17**

2.3 Error Metrics **18**

2.3.1 Mean Absolute Error (MAE) 18

2.3.2 Root mean square error (RMSE) 18
2.3.3 Mean Square Error (MSE) 18
2.3.4 Coefficient of Determination (R2) 19
CHAPTER 3 21
WIND SPEED FORECASTING 21
3.1 Impact of Data Preprocessing on Wind Speed Forecasting 21
3.1.2 Flowchart of Methodology followed 22
3.1.2 Without Data Preprocessing 23
3.1.3 With Data Preprocessing 24
3.2 Results and discussion 25
3.2.1 Case I: Without Data Preprocessing 25
3.2.2 Case II: With data preprocessing 26
3.2.3 Case III: Seasonal Analysis of Wind Energy with Data Preprocessing 28
3.3 Deep Learning Based Wind Energy Forecasting 29
3.3.1 Flowchart of the proposed Methodology 29
3.3.2 Error Metrics of Wind Speed Forecasting 31
3.3.3 Scatter Graph of Proposed Models for Wind Speed Forecasting 31
CHAPTER 4 34
SOLAR IRRADIANCE FORECASTING 34
4.1 Flowchart for Processing of Proposed Model 34
4.2 Error Metrics for Solar Irradiance Forecasting 36
4.3 Scatter graph for the proposed methodology 36
CHAPTER 5 40
ENERGY MANAGEMENT IN ISOLATED MICROGRID 40
5.1 Energy Management Through Energy Vector 40
5.2 Load Demand forecasting 40
5.2.1 Why load forecasting is necessary 41
5.2.2 Flowchart of Load Demand Forecasting and battery for energy management 41
5.2.3 Error Metrics for Load Demand Forecasting 44

5.2.4 Scatter graph of Load Demand Forecasting.....44

5.3 Hybrid Renewable Energy System Integrating Solar, Wind and Battery Storage 46

5.3.1 Graphical representation of solar, wind generation with load demand profile for 24 hrs. 47

5.4 State of Charge of Battery Energy Storage for Three Initial SOC Condition....49

5.4.1 Case 1: Soc for 0.25.....50

5.4.2 Case 2: Soc for 0.55.....50

5.4.3 Case 3: Soc for 0.95.....51

CHAPTER 6.....54

CONCLUSION AND FUTURE SCOPE.....54

6.1 Conclusion 54

6.2 Future Scope.....56

REFERENCES..... 58

Chapter 1

INTRODUCTION

CHAPTER 1

INTRODUCTION

1.1 Background

Renewable energy, particularly wind energy, has gained significant attention globally due to its sustainable nature and minimal environmental impact. The quality and consistency of wind-generated electricity are highly dependent on wind speed. The inherent variability and unpredictability of wind speed can lead to substantial fluctuations and instabilities in power generation and distribution, necessitating accurate wind speed forecasting to optimize the operation, maintenance, and regulation of micro-grids [1]. The intermittent nature of wind energy presents several challenges to power systems, such as difficulties in planning and operations, power quality issues, and the complexity of interconnections. For instance, system operators must allocate additional energy reserves to accommodate discrepancies between scheduled and actual wind power generation. This requirement for extra reserves leads to higher operational costs, which ultimately drives up the cost of electricity [2]. A potential solution to address these challenges is wind energy forecasting. Accurate forecasting can support the development of effective incentives and facilitate efficient day-ahead and hour-ahead electricity markets. Reliable wind energy forecasts enable system operators to better integrate wind power into the grid with fewer complications. Research indicates that precise wind speed and power forecasting play a crucial role in various wind energy applications, ranging from site selection to seamless grid integration [3,4].

Wind energy has rapidly expanded worldwide, supplying a substantial portion of electricity in countries like Denmark. However, the unpredictable nature of wind power introduces more uncertainties compared to conventional energy sources. Traditionally, research has focused on point forecasting methods, but due to the chaotic nature of weather systems, errors are unavoidable. Probabilistic forecasting methods have recently gained popularity as they provide prediction intervals that quantify uncertainty and help in decision-making processes such as optimal operation reserve determination, system security assessment, economic scheduling, and electricity market trading [5]. With the increasing penetration

of wind power into power grids, understanding future uncertainties is crucial for operational strategies, including reserve determination, scheduling, and dispatch [6,7]. Fossil fuels dominate global energy use but release harmful greenhouse gases. Transitioning to clean, renewable energy like wind and solar is essential for meeting rising energy demands while reducing environmental impact [8]. The extensive use of fossil fuels has resulted in severe heatwaves, air pollution, and significant health issues. Moreover, with the rising population, fossil fuel reserves are rapidly depleting [9]. Renewable energy sources, particularly wind and solar power, offer a viable and sustainable alternative due to their renewability and long-term economic benefits [10]. As a result, governments are actively formulating strategies to integrate these energy sources into their power systems. For instance, Germany aims to achieve a fully renewable electricity supply by 2050[11]. Additionally, projections suggest that wind and solar energy will contribute up to 70% of the world's total renewable energy generation [12]. The advantages of renewable energy sources are numerous. They play a crucial role in mitigating global warming, enhancing public health, stabilizing energy costs, and improving energy reliability and resilience [13]. These sources generate power with minimal greenhouse gas emissions, diversify the energy supply, reduce dependence on imported fuels, and contribute to economic growth by creating job opportunities in the manufacturing sector. Accurately forecasting renewable energy consumption and demand is crucial, as it provides valuable insights for energy suppliers' decision-making processes and serves as a strategic guide for governments in developing renewable energy markets. Effective forecasting helps optimize infrastructure development, lower production and manufacturing costs, balance market supply and demand, and minimize resource waste caused by supply-demand mismatches [14]. Machine learning and deep learning techniques have been successfully applied across various domains, including image processing, natural language processing, time series forecasting, and energy load prediction [15]. These models predict wind speed and solar irradiance at a specific location for a chosen future time frame the generation of energy from wind and solar sources is highly dependent on local weather conditions, including temperature, wind speed, air pressure, humidity, and sunlight, as well as their variations. Consequently, predicting and controlling wind and solar power output is challenging due to the constant fluctuations in weather. This poses a significant obstacle to integrating wind and solar energy into power grids, particularly isolated ones. [16,17]

Electricity networks have expanded rapidly during the previous few decades, owing to a significant rise in electrical demands [18]. These loads have increased dramatically, with new load types appearing that necessitate specialized research. As underlined in the Fifth Assessment Report [19,20], several environmental restrictions have been enacted in recent decades to mitigate the effects of climate change. In this context, precise Electric Load Forecasting (ELF) is essential for effective energy distribution and integration of renewable resources, ensuring grid stability, sustainability, and regulatory compliance [21].

With the increasing focus on integrating renewable energy, accurate ELF has become even more vital for efficient energy management and optimizing operations. Advanced machine learning (ML) and real-time data analysis enable facility managers to make wise, data-based decisions taking into account advanced factors such as climate change, seasonality, and evolving user behaviour patterns [22]. Moreover, precise ELF plays a vital role in supporting policymakers and businesses in developing energy strategies that align with global net-zero carbon goals, encouraging energy-efficient behaviour, and driving eco-friendly solutions during smart city and digitalization times.

In recent years, much research has focused on developing more accurate models for ELF, which can be grouped into broad categories of statistical learning (SL), ML, and deep learning (DL) approaches [23,24]. SL models generally use past load data to recognize patterns and correlations through mathematical and statistical methods. Where efficient in particular circumstances, models used often are unsuccessful at predicting complex, non-linear connections typical to energy use and hence confine the level of their prediction accurateness. For alleviation of such factors, ML driven solutions became very important due to the model's natural superiority at tracing advanced, non-linear relations with loads data [25]. With the fast evolution of algorithms and computing power, these techniques have become more effective, particularly when combined with real-time data.

With the rapid escalation of global energy demand and the escalating environmental concerns associated with fossil fuel combustion, the integration of renewable energy technologies into modern power systems has become imperative. Among the various clean energy alternatives, solar photovoltaic (PV) and wind energy conversion systems (WECS) have gained prominence due to their vast availability, scalability, and economic viability.

These resources are pivotal to the global energy transition, driving the decarbonization of the energy sector and supporting the goals of environmental sustainability, energy independence, and long-term economic development.

In the current energy landscape, the decentralization and digitalization of power systems are fundamentally reshaping energy generation, transmission, and distribution. Renewable energy sources, particularly solar and wind, are inherently non-dispatchable and intermittent, introducing significant variability into the grid. As a result, conventional energy management strategies are no longer sufficient to maintain grid reliability and operational efficiency. This necessitates the deployment of advanced forecasting techniques and intelligent energy management systems (EMS) to ensure optimal resource utilization and grid stability.

Today's generation is facing the dual challenge of meeting rising energy consumption while mitigating the adverse impacts of climate change. The electrification of sectors such as transportation, manufacturing, and agriculture further amplifies the need for clean, reliable, and cost-effective electricity generation. Solar and wind energy, owing to their low marginal costs and zero carbon emissions, are key enablers in achieving these objectives.

However, their integration into the power grid introduces technical complexities due to their dependence on meteorological conditions. This results in uncertain power outputs, which can cause voltage fluctuations, frequency deviations, and power quality issues. To address these challenges, accurate and timely renewable energy forecasting becomes essential. By predicting solar irradiance and wind speed with higher precision, system operators can make informed decisions regarding unit commitment, economic dispatch, and real-time grid balancing.

Simultaneously, load forecasting the estimation of future power demand based on historical data and influencing variables is a critical function in energy planning and grid operation. It aids in demand-side management, peak load shaving, and infrastructure development. The integration of short-term, medium-term, and long-term forecasting models enables utilities to align generation schedules with consumption patterns, thereby enhancing the efficiency and resilience of the power system.

1.1.1 Need for Solar and Wind Forecasting with Load Prediction:

The co-optimization of renewable power forecasting and load demand prediction is essential in facilitating the large-scale penetration of solar and wind energy. These forecasts empower system operators to anticipate energy surpluses or deficits and accordingly control Battery Energy Storage Systems (BESS), Flexible AC Transmission Systems, and Demand Response mechanisms. This synergy significantly contributes to the dynamic stability of the grid, reducing reliance on spinning reserves and minimizing system operational costs.

Emerging techniques such as Machine Learning (ML), Artificial Neural Networks (ANN), Support Vector Machines (SVM), and hybrid models are being extensively explored for enhancing forecast accuracy. These data-driven models process vast amounts of real-time data to provide adaptive and robust forecasting capabilities, essential for smart grid operation and renewable energy integration.

As the energy sector progresses toward a low-carbon economy, the importance of integrated forecasting frameworks will continue to grow. The deployment of Internet of Things IoT devices, advanced metering infrastructure, and cloud-based data analytics can further improve the visibility and controllability of distributed energy resources. The convergence of renewable generation, energy storage, and predictive analytics will be the foundation of next-generation power systems. The deployment of solar and wind energy, when complemented with accurate load and generation forecasting, provides a strategic pathway toward sustainable, reliable, and economically efficient power systems. This approach not only addresses the immediate operational challenges but also contributes significantly to the long-term vision of a smart, resilient, and decarbonized energy future. To address this issue, enhancing the accuracy of one-day-ahead forecasting for wind speed and solar irradiation is crucial in reducing uncertainty regarding the availability of renewable energy in different grid operational scenarios. Since solar irradiation directly influences photovoltaic power generation and wind speed affects wind turbine output, it is essential to develop computational models capable of reliably predicting solar irradiation and wind speed over medium- and short-term time frames.

1.2 Problem Statement

The focus of this research work is to ensure a stable, efficient, and sustainable power supply for the isolated commercial area.

In the southern region of Goa, commercial areas operating in isolated systems face challenges in maintaining a reliable and uninterrupted power supply due to the intermittent nature of renewable energy sources like solar and wind. The variability in generation, coupled with fluctuating load demand, leads to difficulties in balancing energy supply and demand. Without proper forecasting and an effective energy management strategy, there is a risk of power shortages, inefficient energy utilization, and increased reliance on backup systems.

Additionally, the absence of an optimally sized Battery Energy Storage System (BESS) further limits the system's ability to store excess energy and supply critical loads during periods of low renewable generation. The lack of insight into the battery's State of Charge (SOC) variations can also impact the system's performance and battery lifespan.

Therefore, there is a need to develop a robust framework for forecasting renewable energy generation and load demand, managing energy flow effectively, and determining the optimal BESS capacity while monitoring SOC variations.

1.3 Objectives of the Study

The primary objectives of this research are:

1. To analyse impact of data preprocessing in renewable energy forecasting.
2. Comparison of machine learning and deep learning model for wind speed and solar irradiance forecasting for enhancing the accuracy.
3. Forecast with the best machine learning model of wind speed, solar irradiance and load demand forecasting with the enhanced accuracy.
4. Manage the energy supply to reliably meet the load demand of a commercial area operating in an isolated power system.
5. Determine the optimal capacity of the Battery Energy Storage System (BESS) required to ensure continuous power supply.
6. Analyse the fluctuations in the State of Charge (SOC) of the battery over time.

1.4 Scope of the Study

This study focuses on the forecasting and management of renewable energy resources and load demand in the southern region of Goa for a commercial area operating in an isolated system. The key aspects covered in the scope include:

1. Renewable Energy Forecasting
2. Load Forecasting
3. Energy Supply Management.
4. Battery Energy Storage System (BESS)
5. State of Charge (SOC) Monitoring

Energy management serves as the central control system for renewable energy setups, coordinating generation, storage, and consumption to enhance efficiency, ensure reliability, and support long-term sustainability.

1.5 Significance of the Study

This research work is significant as it enhances the efficiency and reliability of power generation in isolated systems through intelligent energy management and optimal integration of renewable energy sources. By accurately forecasting solar irradiance, wind speed, and load demand, it ensures better utilization of renewable resources, reducing dependency on conventional energy and lowering environmental impact. The analysis and design of an appropriate Battery Energy Storage System (BESS), along with monitoring the variation in State of Charge (SOC), support efficient energy storage and stable power supply, particularly for commercial areas. This contributes to the development of decentralized and sustainable power systems, promoting energy access in remote regions. Additionally, the study aligns with the goals of smart grid and microgrid advancements by demonstrating how coordinated energy generation, storage, and consumption can lead to improved operational efficiency, reduced carbon emissions, and long-term economic benefits.

1.6 Thesis Organization

This thesis is organized into six chapters:

Chapter2: Overview of Various Machine Learning Tools This chapter reviews brief

description of machine learning tools which are implemented in this work.

Chapter 3: Wind Speed Forecasting: This chapter covers the wind speed forecasting with data preprocessing and without data preprocessing for the Nagpur region and also forecasted wind speed for southern region of Goa.

Chapter4: Solar Insolation Forecasting: This chapter covers the Global Horizontal Irradiance forecasting for commercial load of Southern region of Goa.

Chapter 5: Energy Management in Isolated Microgrid: This chapter details the hybrid renewable energy and load forecasting for energy management of an isolated system and determining the size of battery according to the generation and demand to meet the maximum load demand and examine the fluctuations of state of charge of battery.

Chapter 6: Conclusion and Future Work: This chapter presents the conclusion and future scope of the research work.

CHAPTER 2

BRIEF DESCRIPTION OF VARIOUS MACHINE LEARNING TOOLS

CHAPTER 2

BRIEF DESCRIPTION OF VARIOUS MACHINE LEARNING TOOLS

2.1 Overview of Various machine learning tools

In recent years, the integration of renewable energy sources such as solar and wind into power systems has gained significant momentum due to increasing environmental concerns and the need for sustainable energy solutions. However, the inherent variability and intermittency of renewable energy pose challenges for maintaining a stable and reliable power supply. In this context, machine learning (ML) tools have emerged as powerful solutions for improving the accuracy and efficiency of renewable energy forecasting.

Machine learning refers to a subset of artificial intelligence (AI) that enables systems to learn patterns from historical data and make data-driven predictions or decisions without being explicitly programmed. ML tools can handle large volumes of complex, non-linear, and noisy data characteristics that are common in renewable energy systems. These tools help in modeling the behavior of renewable energy sources based on past weather patterns, solar irradiance, wind speed, temperature, and other relevant environmental factors.

2.1.1 Multiple Linear Regression (MLR)

Regression analysis is a statistical approach used to determine the relationship between variables when a cause-and effect dynamic is present. The main purpose of univariate regression is to explore how a single independent variable influences a dependent variable and to create a linear equation that describes their relationship. This technique enables analysts to determine the variation of the model and the relative contribution of each independent variable in the total variance. When the model involves one dependent variable and multiple independent variables, it is referred to as MLR. The MLR model is represented as follows:

$$y = \alpha_0 + \alpha_1 A_1 + \alpha_2 A_2 + \cdots + \alpha_n A_n \qquad (2.1)$$

where, y is a dependent variable and A_1, A_2, ...An are independent variables. α_0 ,α_1 , α_2, ..., α_n are the coefficients, representing the intercept and slopes of the independent

variables respectively. MLR is utilized to enhance the prediction process for wind speed and power output by examining the effects of various meteorological and environmental variables. It establishes a mathematical connection between multiple factors, including temperature, pressure, wind direction, and historical wind speeds, with the target output of future wind speed or energy generation. By pinpointing significant variables and assessing their effects, MLR can refine forecasting models, minimize prediction errors, and increase the dependability of integrating wind energy into the power grid. This optimization improves operational efficiency, lowers costs related to energy discrepancies, and facilitates more effective planning and decision-making in the management of wind energy:

2.1.2 Random Forest Regression (RFR)

A random forest regression model combines multiple decision trees into a single ensemble. Each tree is built from a different subset of the data and makes an independent prediction. The final output is determined by averaging or taking a weighted average of all the trees' predictions Random Forest is a supervised learning algorithm that uses an ensemble of decision trees to predict outcomes. It operates by creating multiple decision trees from random subsets of the data using a technique called bagging. Each tree makes its own prediction, and the model aggregates these predictions to generate a final result. In regression tasks, random forest predicts continuous target variables, reducing variance and improving accuracy by combining the outputs of several trees. This approach allows it to handle larger datasets and capture more complex relationships than individual decision trees.

Model Predictive Control (MPC), on the other hand, utilizes a mathematical model of the motor system to predict future states and optimize control actions. This approach proves invaluable in real-time optimization scenarios, where the controller can anticipate and counteract disturbances before they affect the system's performance. MPC's ability to handle multiple constraints and optimize performance in a dynamic environment makes it particularly suitable for high-precision applications.

2.1.3 Decision Tree (DT)

A decision tree is a non-parametric, supervised learning algorithm used for both classification and regression problems. It is represented as a tree structure where each

internal node represents a test on an attribute, each branch represents the outcome of the test, and each leaf node represents a class label or a predicted value. The goal of a decision tree is to split the dataset into subsets based on the value of an attribute, repeating this process until each subset contains only instances that belong to a single class or have similar values. The method is applied recursively to each subset until a stopping condition is met, such as a maximum tree depth. Variance is calculated as the mean of the squared differences between the output values and their mean. DT handle multi-output problems with minimal data preparation, and their structure resembles a map of potential outcomes based on a series of decisions. The model learns to predict a target value by forming simple decision rules from the dataset's features. Their ability to model complex, non-linear relationships and handle both categorical and numerical data makes them versatile across various industries. While decision trees are prone to overfitting, techniques like pruning help mitigate this issue, ensuring that the model generalizes well to unseen data. With the right balance of depth, pruning, and attribute selection, decision trees provide a robust and transparent way to solve machine learning problems.

2.1.4 Support Vector Regression (SVR)

SVR is a type of support vector machine (SVM) that is used for regression tasks. It tries to find a function that best predicts the continuous output value for a given input value**,** SVR can use both linear and non-linear kernels. A linear kernel is a simple dot product between two input vectors, while a non-linear kernel is a more complex function that can capture more intricate patterns in the data. The choice of kernel depends on the data's characteristics and the task's complexity application. SVR is a ML technique that employs the principle of structural risk minimization to classify, regress, and perform other learning tasks using a limited number of training samples. Initially proposed by Vapnik for identifying subtle patterns in complex datasets, SVR formulates the statistical problem using quadratic programming. The main idea behind SVR is to separate data points into two classes by identifying a hyperplane that maximizes the margin between them. SVR, a variant of SVR, aims to minimize generalization error and relies on a limited subset of training data. By projecting data into a higher-dimensional space, SVR effectively reduces the error margin in regression, making it suitable for forecasting.

In SVR, the cost function disregards any training data within a specified margin (controlled by the hyperparameter epsilon) around the model's predictions, allowing for a balance between bias and variance. The SVR forecasts are compared against a persistence forecast, utilizing a moving time window where the training data is shifted for each prediction step. The initial training of the SVR algorithm is conducted in Python using a dataset containing past data of the target parameter. The hyperparameters of SVR are tuned by using Bayesian search. the values for 'C' and 'gamma' are 11.2741391 and 0.520 respectively. This process of adding various features to improve forecast accuracy is known as feature selection or feature tailoring. For the given data variables A and y, the equation of SVR function is given by Eq (2.2)

$$\boldsymbol{y = WA + Z} \tag{2.2}$$

where, y a dependent variable, A is independent variable, W is weight vector and Z is bias term.

2.1.5 Convolutional Neural Network (CNN)

CNNs are a powerful class of neural networks designed to filter and extract features from input data. Widely used in image processing, classification, and computer vision, they also extend to NLP, audio recognition, medical applications, and load forecasting [26]. However, variations in load profiles due to human behavior, time, and weather add complexity. CNNs can be used for forecasting by leveraging their ability to extract features and patterns from time series data. By processing time series data through convolutional layers, they can automatically learn relevant features that improve forecasting accuracy. The output of a convolutional layer can be mathematically represented as:

$$\boldsymbol{Y[i,j] = \sum_{m=0}^{k-1} \sum_{n=0}^{k-1} X[i+m, j+n] \cdot W[m,n]} \tag{2.3}$$

`Where:

- X is the input time series data.
- W is the weight matrix of the convolutional filter.
- b is the bias term.
- Y is the output feature map.
- k is the size of the filter

- i and j are the indices of the output feature map

One-dimensional CNNs (1D-CNNs) are commonly used for time series forecasting due to their ability to directly process sequential data. The 1D-CNN architecture consists of convolutional layers, pooling layers, and fully connected layers. The convolutional layers extract temporal features, pooling layers reduce the dimensionality of the feature maps, and fully connected layers make the final predictions. After feature extraction, the CNN model makes predictions based on the learned features. For example, a fully connected layer can be used to map the extracted features to the predicted values. The prediction can be mathematically represented as

$$\hat{\boldsymbol{y}} = \boldsymbol{Activation\ function}\left(\boldsymbol{W_{fc}} \cdot \boldsymbol{Y} + \boldsymbol{b_{fc}}\right) \qquad (2.4)$$

Where:

- $\hat{\boldsymbol{y}}$ is the predicted value
- $\boldsymbol{W_{fc}}$ is the weight matrix of the fully connected layer
- $\boldsymbol{Y}$ is the output of the previous layer
- $\boldsymbol{b_{fc}}$ is the bias term of the fully connected layer
- Activation Function is a non-linear activation function

Eq (8) provide a basic framework for understanding how CNN's can be applied to time series forecasting, highlighting the key steps of feature extraction and prediction.

2.1.6 XGBOOST

XG Boost is an advanced decision tree-based optimization algorithm that enhances the gradient descent method. It leverages gradient descent to optimize the loss function while incorporating regularization parameters to mitigate overfitting [27]. The core principle of the XG Boost algorithm is to minimize an objective function that comprises both the loss function and regularization terms

$$\mathbf{L^{(t)}} = \sum_{\mathbf{i=1}}^{\mathbf{n}} \mathbf{l}\left(\mathbf{y_i}, \hat{\mathbf{y}}_{\mathbf{i}}^{(\mathbf{t-1})} + \mathbf{f_{t(x_i)}}\right) + \boldsymbol{\Omega}(\mathbf{f_t}) \qquad (2.5)$$

Where, l is the loss function that represents the error between observed data yi and predicted data $\hat{y}_i$, f_t is the model of the t^{th} tree, and t is the iteration index during the optimization process. The detail of regulation term $\Omega(f_t)$ can be expressed as [28,29]

$$\boldsymbol{\Omega(f) = \gamma T + \frac{1}{2}\lambda||w||^2} \quad (2.6)$$

Eq (2.6) represents a regularization function used to control model complexity and prevent overfitting in ML. The first term, γT gamma, promotes sparsity by encouraging some parameters to be zero, making the model more interpretable and reducing unnecessary complexity. The second term $\frac{1}{2}\lambda||w||^2$ represents L2 regularization, which penalizes large weight values to ensure stability and better generalization. By combining both L1 sparsity and L2 smoothness regularization, this function is commonly used in techniques like Elastic Net, balancing feature selection and model robustness for improved performance.

2.1.7 Long Short-Term Memory (LSTM)

LSTM models, a type of RNN, are highly effective at capturing and preserving long-term dependencies, making them well-suited for applications like financial time series forecasting [30]. The core functionality of an LSTM cell is defined by the following four equations.

The forget gate equation is:

$$\mathbf{f_t = \sigma(W_f * [h_{t-1}, x_t] + b_f)} \quad (2.7)$$

where,W_f represents the weight matrix associated with the forget gate. $[h_{t-1}, x_t]$ denotes the concatenation of the current input and the previous hidden state.$b_f$is the bias term for the forget gate and σ is the sigmoid activation function.

$$\boldsymbol{i_t = \sigma(W_i \cdot [h_{\{t-1\}}, x_t] + b_i)} \quad (2.8)$$

Eq (2.8) determines how much of the new input should be allowed into the memory cell. The sigmoid activation function σ ensures that the values range between 0 and 1, allowing selective passage of information,W_i And b_iAre the weight matrix and bias for the input gate, The input consists of the previous hidden state $[h_{\{t-1\}}]$And the current input xt . If it is close to 1, more information is passed into the cell state; if it's close to 0, less information is retained.

$$\boldsymbol{c_t = \tanh(W_c \cdot [h_{\{t-1\}}, x_t] + b_c)} \quad (2.9)$$

Eq (2.9) calculates the candidate cell state, which represents new information that could be added to the memory. he tanh activation function scales values between -1 and 1, helping to regulate updates. Wc and bc are the weight matrix and bias for this operation. The combination of previous hidden state $\left[h_{\{t-1\}}\right]$ and current input xt helps decide what new information is relevant for updating the cell state.

The equation for the output gate is:

$$\boldsymbol{o\_t = \sigma(W_o \cdot \left[h_{\{t-1\}}, x_t\right] + b_o} \qquad (2.10)$$

In Eq (2.10) The sigmoid activation function σ ensures that the output values range between 0 and 1, allowing the model to selectively filter information. The term W_o is the weight matrix, and b_o is the bias for the output gate. The input to this function consists of the previous hidden state $h_{\{t-1\}}$ and the current input x_t , both of which are concatenated before being processed. A value close to 1 means that more information from the cell state will be passed to the next layer, while a value close to 0 restricts the amount of information that is carried forward. This mechanism helps the LSTM model focus on relevant information and discard unnecessary details, improving its ability to handle long-term dependencies in sequential data.

2.2 Implementation of Machine Learning Model

- **Define the Problem**
  Identify the task
- **Collect the Data**
  Get relevant data from files, APIs, or databases.
- **Load & Explore the Data**
  Use pandas and plots to understand the data.
- **Preprocess the Data**
  Clean, encode, scale, and split into train/test sets.
- **Select a Model**
  Choose a model based on task
- **Train the Model**
  Fit the model using training data.

- **Make Predictions**

  Use the model to predict on test data.
- **Evaluate the model Performance**

  Use error metrics

2.3 Error Metrics

Error metrics are quantitative measures used to evaluate how well a machine learning model performs on unseen data. They compare the model's predicted outputs to the actual values.

2.3.1 Mean Absolute Error (MAE)

MAE calculates the average magnitude of the difference between observed values and predicted values, without considering the direction of the errors.[31] The formula for MAE is expressed as Eq (2.11)

$$MAE = \frac{1}{T}\sum_{i=1}^{T}|y_\alpha - \hat{y}_\alpha| \tag{2.11}$$

where T is the number of samples, y_α is the observation and is the prediction.

2.3.2 Root mean square error (RMSE)

RMSE calculates the square root of the average of the squared differences between observed and predicted values [31]. RMSE can be defined as Eq (2.12)

$$RMSE = \sqrt{\frac{1}{T}\sum_{i=1}^{T}(y_\alpha - \hat{y}_\alpha)^2} \tag{2.12}$$

2.3.3 Mean Square Error (MSE)

MSE measures the average of the squared differences between the actual values in the test set and the predicted values [32]. The formula to calculate MSE is given by Eq (2.13)

$$MSE = \frac{\sum_{k=1}^{T}\left(y_{test} - y_{predk}\right)^2}{n} \tag{2.13}$$

where, y_{test} is actual value for kth observation, y_{predk} is predicted value for k^{th} observation.

2.3.4 Coefficient of Determination (R2)

R^2 is a statistical metric that indicates how well the predictions align with the observed data. It evaluates the ratio of the residual sum of squares to the total sum of squares of deviations [32,33].

R^2 is given as Eq (2.13)

$$R^2 = 1 - \frac{\sum_{i=1}^{T}(y_\alpha - \hat{y}_\alpha)^2}{\sum_{i=1}^{T}(y_\alpha - \bar{y}_\alpha)^2} \quad (2.14)$$

The symbol represents the average of the observed values. Its value lies between 0 and 1, with values closer to 1 indicating better prediction accuracy.[34]

CHAPTER 3

WIND SPEED FORECASTING

CHAPTER 3

WIND SPEED FORECASTING

Wind speed forecasting is the process of predicting the future speed of wind at a specific location over a certain period using various methods such as physical models, statistical techniques, or machine learning algorithms. It helps estimate how wind conditions will change in the near or long term.

It is necessary because wind speed directly affects several critical sectors. In wind energy generation, accurate forecasting ensures efficient turbine operation, power scheduling, and grid stability. In aviation and marine industries, it supports safe navigation and flight planning. It also aids in weather prediction, disaster management (e.g., predicting storms or cyclones), and agriculture, where wind affects crop spraying and pollination. Without accurate wind forecasts, there can be significant operational, financial, and safety risks across these domains.

3.1 Impact of Data Preprocessing on Wind Speed Forecasting

Data preprocessing has a significant impact on the accuracy and effectiveness of wind speed forecasting, especially when using machine learning or statistical models. Raw wind speed data collected from sensors or weather stations often contains noise, missing values, and inconsistencies that can severely affect the model's performance if not handled properly. Preprocessing involves cleaning the data by filling or removing missing values, filtering out noise, and detecting and eliminating outliers that may result from sensor faults or extreme weather events. This ensures that the forecasting model learns from high-quality, reliable inputs. Additionally, feature engineering—such as including relevant variables like wind direction, temperature, humidity, pressure, and time-based features (e.g., hour of the day, season) adds contextual information that improves the model's understanding of the patterns affecting wind speed. Scaling and normalization of numerical features are also essential, particularly for algorithms sensitive to input magnitudes like Support Vector Machines or Neural Networks, as they ensure faster convergence and balanced model learning. Categorical variables like location or weather conditions must be encoded into numerical formats for the model to process them effectively**.** Data

preprocessing is not just a preliminary step but a foundational process that directly influences the reliability, efficiency, and accuracy of wind speed forecasting.

3.1.2 Flowchart of Methodology followed

Fig 3.1 illustrates a flowchart of comprehensive process for wind speed forecasting using hourly wind data. The methodology begins with the collection of raw data, which can either be used directly without preprocessing or undergo a detailed preprocessing workflow.

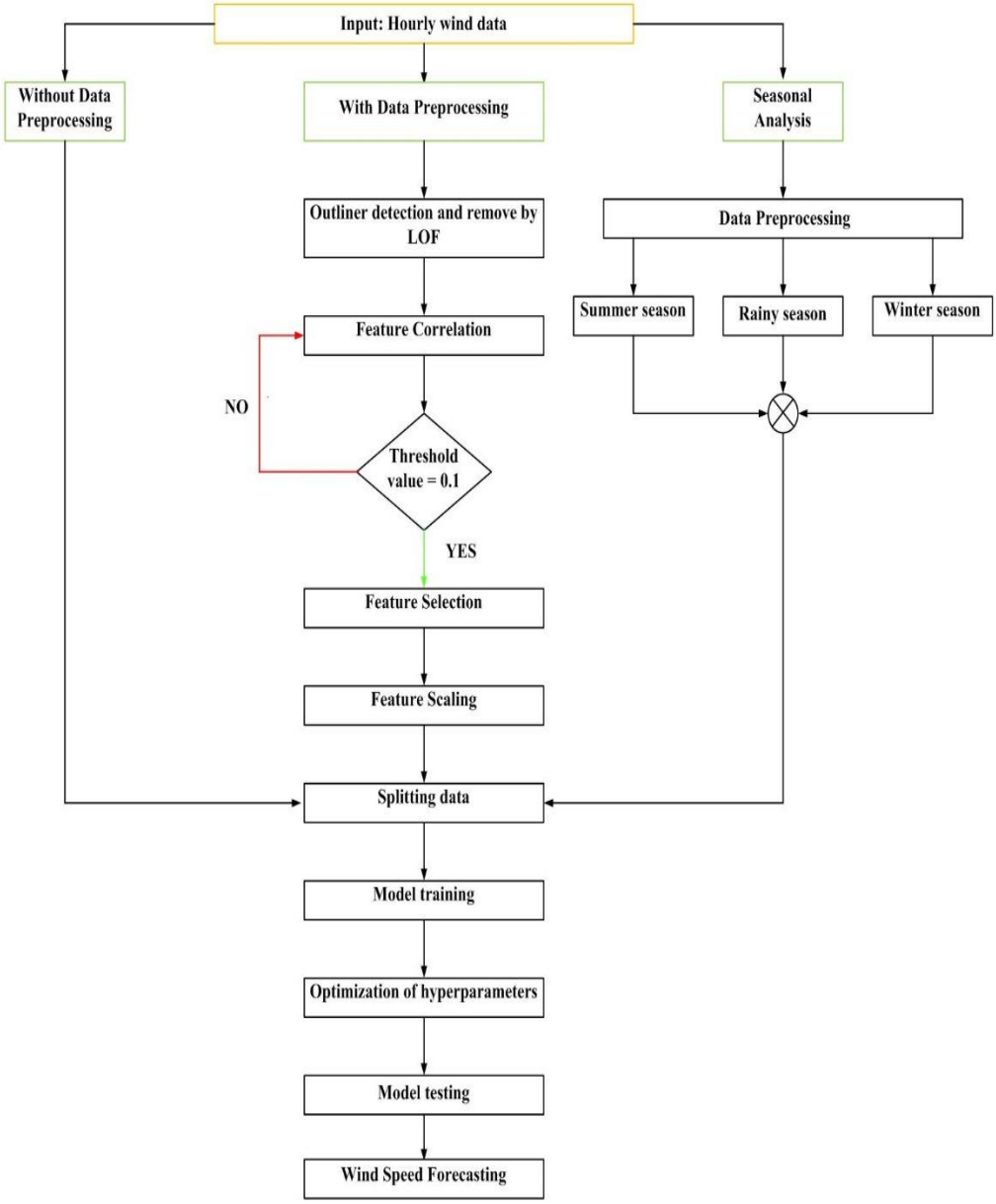


Figure 3.1: Flowchart to Follow Method

In the preprocessing path, outliers are first detected and removed using the Local Outlier Factor (LOF) method to improve data quality. Next, feature correlation is analysed, and only features meeting a specified threshold value (0.1) are selected for further use. Once relevant features are selected, feature scaling is applied to normalize the data. Concurrently, a seasonal analysis is conducted by dividing the dataset into summer, rainy, and winter seasons, each of which is individually pre-processed and then merged. Following preprocessing, the data is split into training and testing sets. Machine learning models are then trained, and their performance is optimized by fine-tuning hyperparameters. Finally, the trained models are tested, and the system is used to forecast wind speed accurately. This structured approach enhances the robustness and accuracy of wind speed prediction by incorporating data cleaning, feature engineering, and seasonal variability.

3.1.2 Without Data Preprocessing

Hourly data was gathered from January 2018, to December 2019, using the National Solar Radiance Database (NSRDB) for the coordinates 17.17 latitude and 75.02 longitude. The dataset includes ten features: date, time, temperature, wind speed, wind direction, precipitable water, relative humidity, Diffuse Horizontal Irradiance (DHI), pressure, and Global Horizontal Irradiance (GHI). In case of without data preprocessing, all the parameters with wind speed forecasting were taken into consideration.

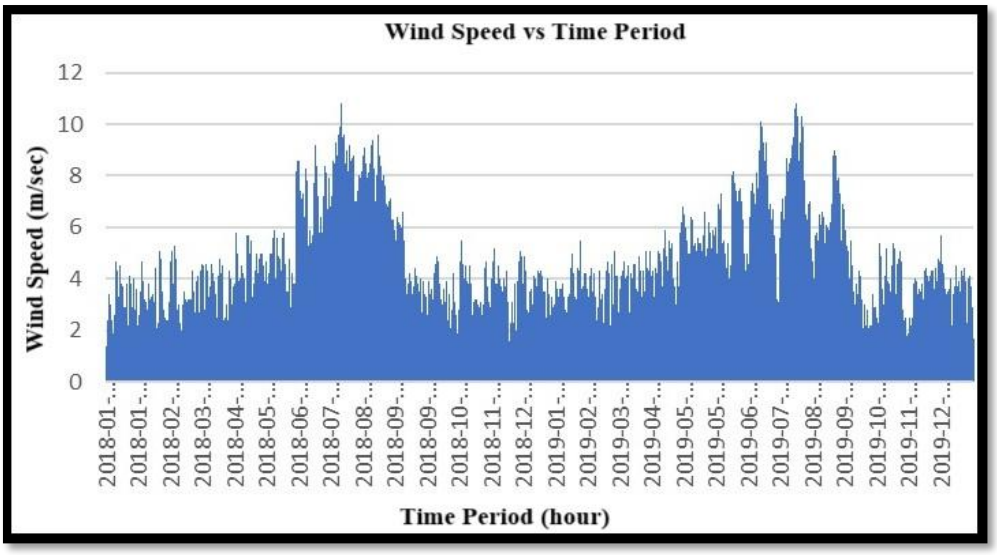


Figure 3.2: Wind speed over the time period

Cleaning the dataset, finding outlier indices, creating a data frame, extracting the outlier frame, do not take place in this process.

The Fig. 2 depicts the wind speed over the duration of time series plot showing wind speed (m/sec) versus time period (in hours) over a span of two years.

3.1.3 With Data Preprocessing

Before making predictions, data preprocessing is a crucial first step. This process involves detecting and removing outliers, which are data points that significantly deviate from the rest of the observations. The Local Outlier Factor (LOF) method is used to identify these outliers which extract numeric data, identifies the outliers, finding outliers indices, extracting the outlier points and cleans the dataset for this we have used correlation method which specifies the weather relation attributes to restrain from the dataset, these attributes include temporal data and various meteorological variables [33].

Initially, there were 17,520 samples, but after outlier removal, 770 samples were eliminated, leaving 16,750 clean samples for further processing. After outlier removal, assessing the correlation between variables is essential to determine the most relevant input features for accurately forecasting wind energy. A correlation matrix explores the relationships between input and output variables. Fig. 3 shows the correlation matrix which is generated using the ML methods.

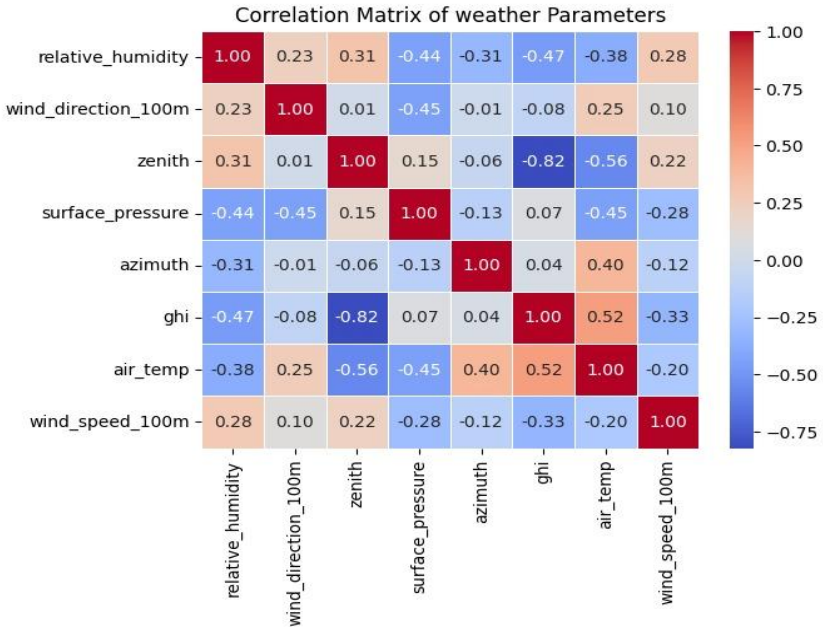


Figure 3.3: Correlation Matrix

The correlation matrix score ranges from -1 to 1, where 1 indicates a positive linear relationship (direct relationship), 0 signifies no linear relationship, and -1 represents a negative relationship (inverse relationship). These scores help identify the true correlations between variables. The correlation matrix, illustrated in Fig. 3.3is used to select four key input features for forecasting: temperature, wind speed, relative humidity, and DHI.
The dataset is then split into two parts: X, which includes the independent variables, and Y, which represents the dependent variable. In this case, X consists of the selected features temperature, solar irradiance, relative humidity, and DHI while Y corresponds to wind speed.

3.2 Results and discussion

3.2.1 Case I: Without Data Preprocessing

In this work wind energy has been forecasted with the help of different types of ML techniques on three different case studies

Table 3.1: Without Data Preprocessing

Methods	**RMSE**	**MSE**	**MAE**	**R-SQUARE**
MLR	5.4985	3.0234	5.1334	0.3459
RFR	5.2569	2.9435	2.9969	0.6634
SVR	3.3618	3.2946	2.8913	0.70109
DT	4.1235	4.0536	3.9837	0.6879

Table 3.1 shows the comparison between these four methods. From the table it is clear that SVR gives better accuracy in this case. Its R-Square is closer to 1 as compared to MLR, RFR and DT.

Fig. 3.4 shows the different results that are obtained when the wind forecasting is done without preprocessing the data. The wind energy is forecasted by using four ML methods MLR, SVR, RFR and DT without data preprocessing of their parameters and models are evaluated by comparing their MAE, MSE RMSE and R-Square.

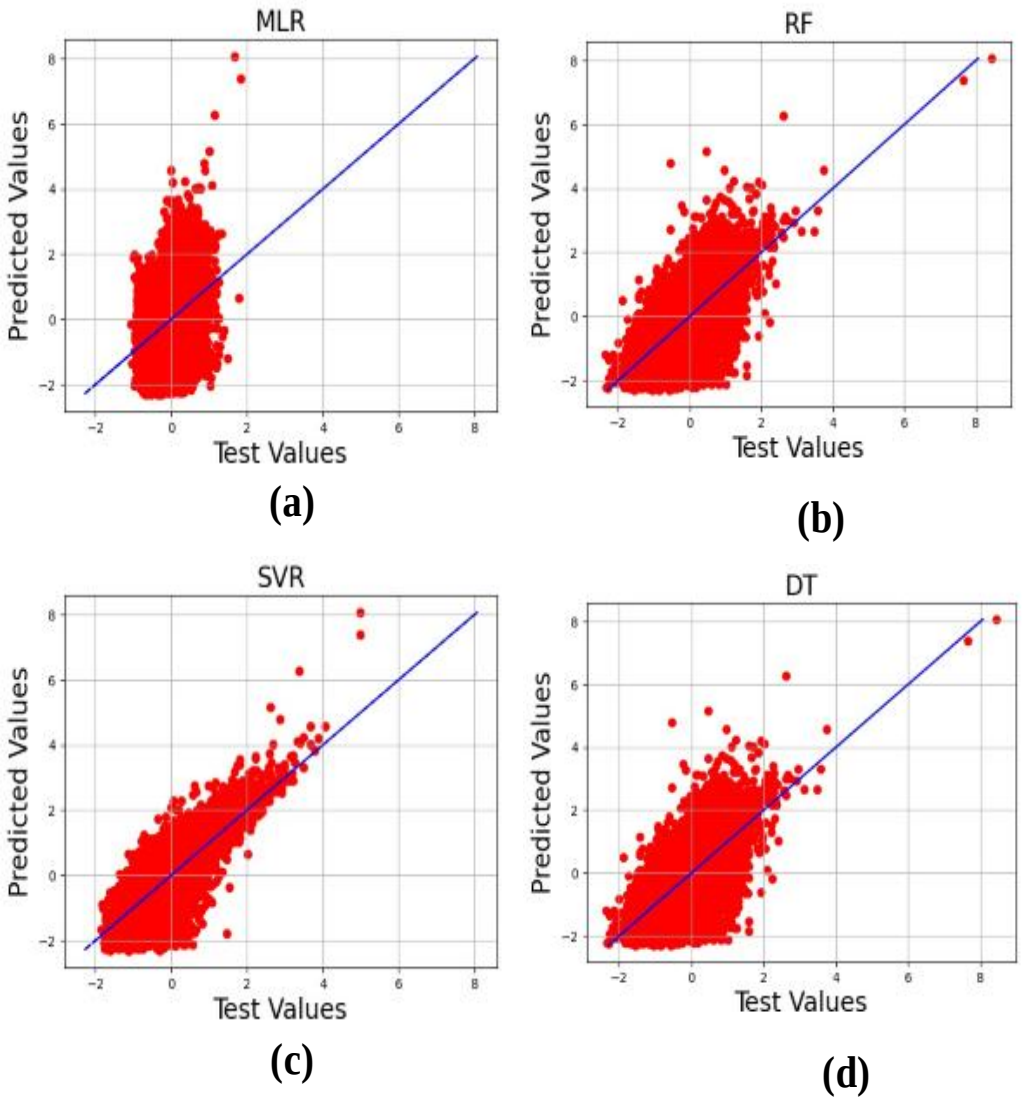


Figure 3.4: Scatter graph for wind speed forecasting for case I (a) MLR, (b) RFR (c) SVR and (d) DT

3.2.2 Case II: With data preprocessing

In this case also like earlier four ML models i.e., MLR, SVR, RFR and DT are trained. After data preprocessing their parameters and test results are evaluated by using metrics RMSE, MSE, MAE and R-Square. Below Table 3.2 shows the comparison of results of MLR, SVR, RFR and DT after data preprocessing. The Bayesian search algorithm and grid search algorithms are used for data preprocessing of MLR, SVR, RFR, and DT.

Table 3.2: With Data Preprocessing

Methods	RMSE	MSE	MAE	R-Square
MLR	0.8144	0.6632	0.6209	0.3846
RFR	0.3882	0.1507	0.2724	0.7489
SVR	0.3241	0.105	0.2301	0.7713
DT	0.4249	0.1805	0.3117	0.6381

Fig. 3.5 depicts the results that are obtained when the data preprocessing is done. From table II it is evident that SVR is the ML technique which gives the higher accuracy in this case.

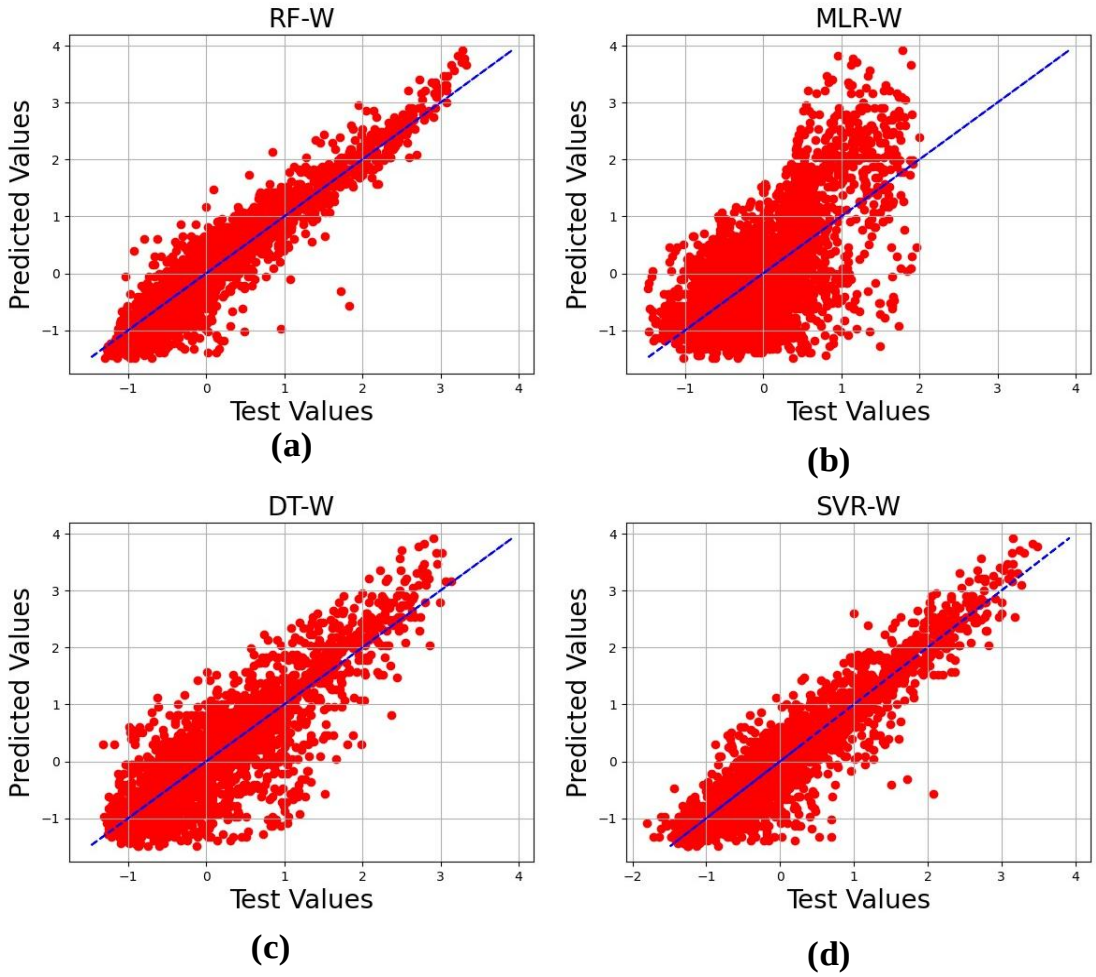


Figure 3.5: Scatter graph for wind speed forecasting for case II (a) MLR, (b) RFR, (c) SVR and (d) DT

The SVR method prove to be efficient when the data preprocessing. After preprocessing, SVR model outperformed in terms of accuracy and showed reduced errors. The comparison between the actual preprocessing and the predicted values for all four methods, both before and after data preprocessing. The graph reveals that the number of outliers decreases significantly after preprocessing, and the results are more refined, indicating that accuracy in prediction can be improved by optimizing the forecasting models' parameters.

3.2.3 Case III: Seasonal Analysis of Wind Energy with Data Preprocessing

This case deals with seasonal analysis of MLR, SVR, RFR and DT. Data from 1stJanuary 2018 to 28th December 2019 is taken for seasonal analysis. Again, correlation between for seasonal data has been calculated and according to the score of correlation matrix input features are selected for prediction of wind energy. In this case models after data preprocessing are used and trained. Test results are evaluated by calculating RMSE, MAE, MSE and R2. Table III shows the comparison of these four methods on the basis of their performance in different seasons.

Table 3.3: Seasonal Analysis with Data Preprocessing

Season	Methods	RMSE	MSE	MAE	R-SQUARE
Summer	**MLR**	0.7783	0.6058	0.5967	0.3846
	RFR	0.4934	0.2434	0.3404	0.7527
	SVR	0.4747	0.2253	0.3278	0.7711
	DT	0.5969	0.3563	0.4245	0.638
Rainy	**MLR**	0.7783	0.6058	0.5967	0.3846
	RFR	0.4962	0.2462	0.3424	0.7498
	SVR	0.4747	0.2253	0.3278	0.7711
	DT	0.5969	0.3563	0.4249	0.6381
Winter	**MLR**	0.7783	0.6058	0.5967	0.3446
	RFR	0.4944	0.2444	0.3413	0.7515
	SVR	0.4747	0.2253	0.3278	0.7711
	DT	0.5969	0.3563	0.4244	0.6381

Table 3.3 represents the comparative error analysis on predicted Wind speed for summer, rainy and winter season with data preprocessing. From table 3.3 it is clear that RF has better prediction in summer, predicted values are closer to actual values as compared to other three methods. It is observed that the R2 value is more in all the three seasons with SVR as compare to MLR, RFR and DT. From the above analysis, it is concluded that SVR provides better accuracy. As SVR gives predictions within an epsilon tube, higher the value of epsilon lesser the model sensitive to errors. However, a critical analysis based on seasonal variation suggests that one technique sometimes predict the most accurate results for various scenarios.

3.3 Deep Learning Based Wind Energy Forecasting

Data-driven models, primarily based on time-series analysis, have lower computational complexity and are well-suited for short-term forecasting. In contrast, physical models use mathematical equations to describe atmospheric dynamics and solar radiation, making them ideal for medium- and long-term predictions. Given their efficiency and accuracy in short-term forecasting, this study focuses on data-driven models.

This study compares XG Boost, CNN, and LSTM for wind speed forecasting, demonstrating that LSTM outperforms other models with the lowest errors and highest accuracy. By leveraging meteorological data from the Sol cast API toolkit, it enhances renewable energy prediction for grid-isolated systems in southern Goa. Forecast wind speed for the southern region of Goa having latitude 15.1177° N and longitude 74.1240° E is analysed over a time period, from January 2019 to December 2023.

3.3.1 Flowchart of the proposed Methodology

Fig. 3.6 flowchart illustrates the process of wind speed prediction using machine learning models. It begins with collecting raw wind speed data, followed by detecting and removing outliers to ensure data quality. The data then undergoes feature scaling through standardization to normalize its range. After preprocessing, the dataset is split into training and testing subsets. Three models XG Boost, CNN, and LSTM are trained on the training data to learn patterns and trends. These models are then tested on the unseen data to evaluate their predictive accuracy. Finally, the predicted results from each model are compared to determine which one performs best in forecasting wind speed.

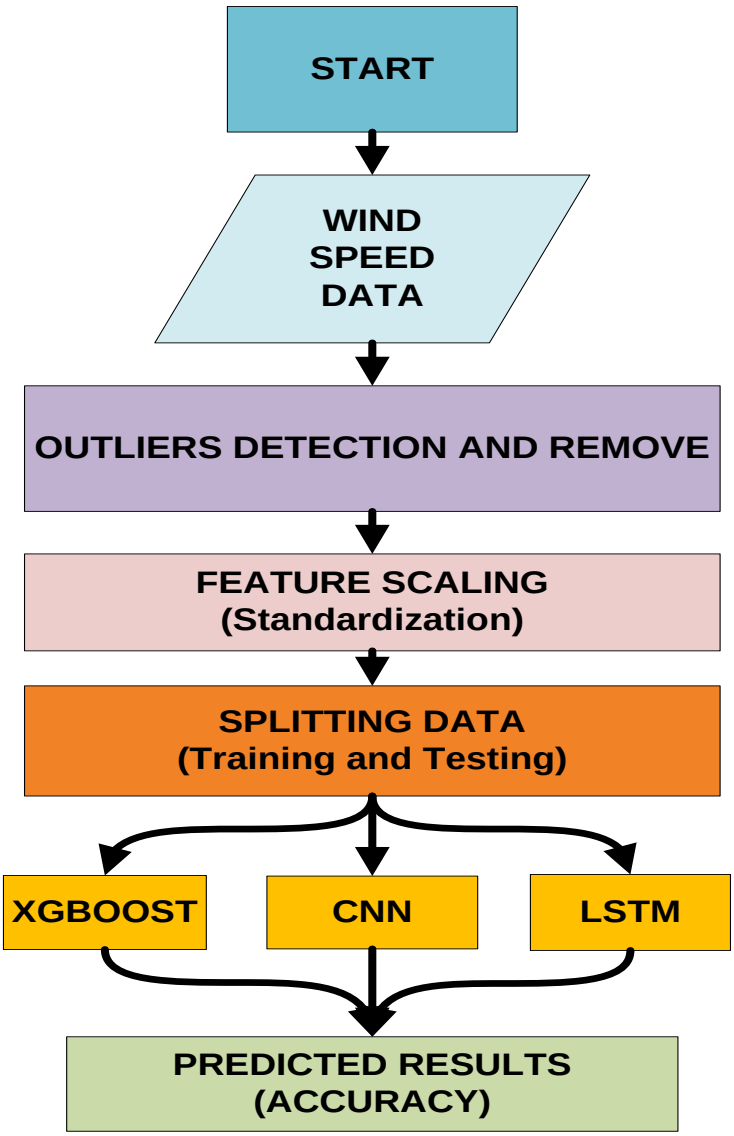


Figure3.6: Flowchart for Processing of Proposed Models

Wind speed forecasting is a crucial task in renewable energy management, aviation, and meteorology, requiring accurate predictions to optimize efficiency and safety. In this analysis, machine learning models such as XGBOOST, LSTM, and CNN were employed to forecast wind speed. Each model was evaluated based on error metrics like RMSE, MSE, MAE, and R^2. According to the results, LSTM demonstrated the best performance, achieving the lowest error values and the highest R^2 score, indicating superior predictive accuracy Compared to XGBOOST and CNN, LSTM effectively minimized errors, making it the most reliable model for wind.

3.3.2 Error Metrics of Wind Speed Forecasting

Table 3.4: Error Metrics of Wind Speed Forecasting

METHODS	RMSE	MSE	MAE	R^2
XGBOOST	0.3756	0.1421	0.2981	0.959
CNN	0.3756	0.2223	0.3734	0.9407
LSTM	0.1793	0.0321	0.1366	0.9684

Table 3.4 compares the performance of XGBOOST, LSTM, and CNN models for wind speed prediction using four key metrics: RMSE, MSE, MAE, and R^2. Among the three, LSTM demonstrates the best performance with the lowest RMSE of 0.1793, MSE of 0.0321, and MAE of 0.1366, indicating higher accuracy and lower prediction errors. Additionally, it achieves the highest R^2 value of 0.9684, showing a strong fit to the data. In contrast, XGBOOST and CNN exhibit similar RMSE of 0.3756 but differ slightly in other metrics, with CNN having a higher R^2 of 0.9407 but also a higher MAE of 0.3734 compared to XGBOOST. Overall, LSTM outperforms both models and appears to be the most reliable choice for wind speed prediction.

3.3.3 Scatter Graph of Proposed Models for Wind Speed Forecasting

Fig3.7 present a comparison between actual and predicted wind speeds in Goa using three different machine learning models: XG Boost, CNN, and LSTM. In the first graph, the XG Boost model shows a strong correlation between predicted and actual wind speeds, with data points closely following the diagonal line representing perfect predictions. However, there is some scattering, especially at higher wind speeds, indicating minor prediction errors. The CNN model's graph also demonstrates a reasonable alignment with the actual values, but with slightly more dispersion, particularly at the extremes, suggesting that CNN has more difficulty accurately predicting higher wind speeds.

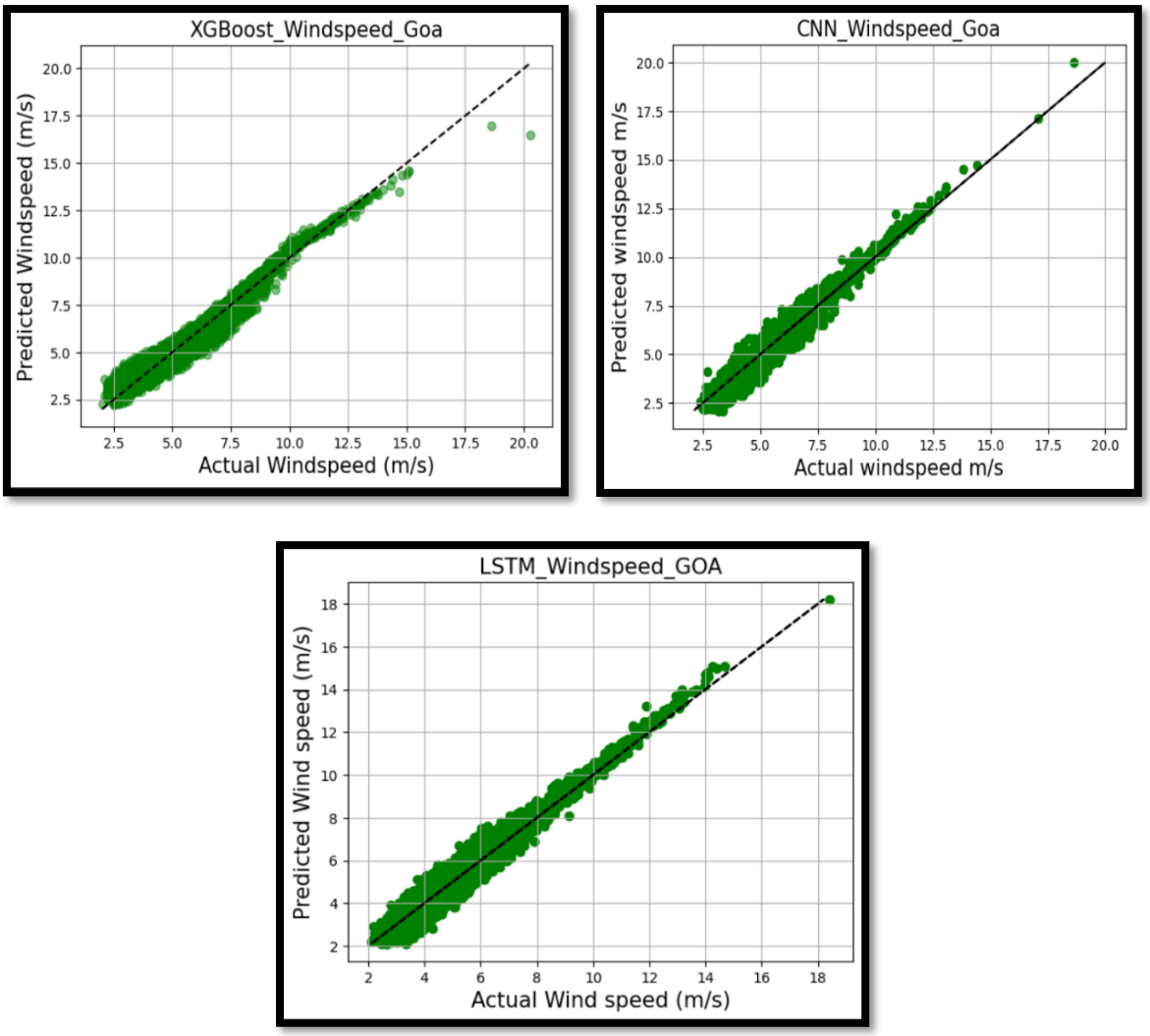


Figure 3.7: Scatter graph for Wind Speed Forecasting for (a)XG BOOST, (b) CNN (c) LSTM

The LSTM model exhibits the tightest clustering of points along the diagonal line, reflecting highly accurate predictions and consistent performance. This indicates that LSTM, with its ability to capture temporal dependencies, outperforms both XG Boost and CNN in forecasting wind speed. Overall, while all models perform well, LSTM provides the most reliable and precise predictions for wind speed in this context.

CHAPTER 4

SOLAR IRRADIANCE FORECASTING

CHAPTER 4

SOLAR IRRADIANCE FORECASTING

In this chapter, we will implement the three machine learning techniques to forecast solar irradiance, Solar Irradiance forecasting is essential for efficient solar energy integration into the power grid. Since solar generation varies due to weather conditions, accurate forecasting helps optimize load balancing, reduce reliance on fossil fuels, and enhance energy storage [35]. This study implements XG Boost, CNN, and LSTM for solar irradiance prediction. XG Boost excels in short-term forecasting but struggles with long-term dependencies. CNN captures spatial and temporal patterns but lacks sequential memory. LSTM outperforms both, delivering higher accuracy and lower errors by effectively learning long-term dependencies in time-series data. Reliable forecasting improves grid stability, optimizes battery storage, and enhances power plant operations. The predicted values from all three models, shown in the graph below, highlight LSTM's superior performance, contributing to a more stable and sustainable energy system. This study compares XG Boost, CNN, and LSTM for solar irradiance forecasting, demonstrating that LSTM outperforms other models with the lowest errors and highest accuracy. By leveraging meteorological data from the Sol cast API toolkit, it enhances renewable energy prediction for grid-isolated systems Forecast solar irradiance for the southern region of Goa having latitude 15.1177° N and longitude 74.1240° E is analyzed over a time period, from January 2019 to December 2023.

4.1 Flowchart for Processing of Proposed Model

Fig 4.1 flowchart illustrates the systematic workflow for forecasting solar energy using machine learning models. The process begins with the collection of solar input data, which may include parameters such as solar irradiance, temperature, and time-based features. Once the data is acquired, it is processed in the "Run up the data" step, which involves loading and structuring the dataset for further analysis. The next stage involves outlier detection, where abnormal or extreme values in the data are identified and addressed to ensure accuracy and consistency. Following this, feature scaling is applied to normalize the data, bringing all features to a common scale, which is crucial for improving the performance of models like CNN and LSTM.

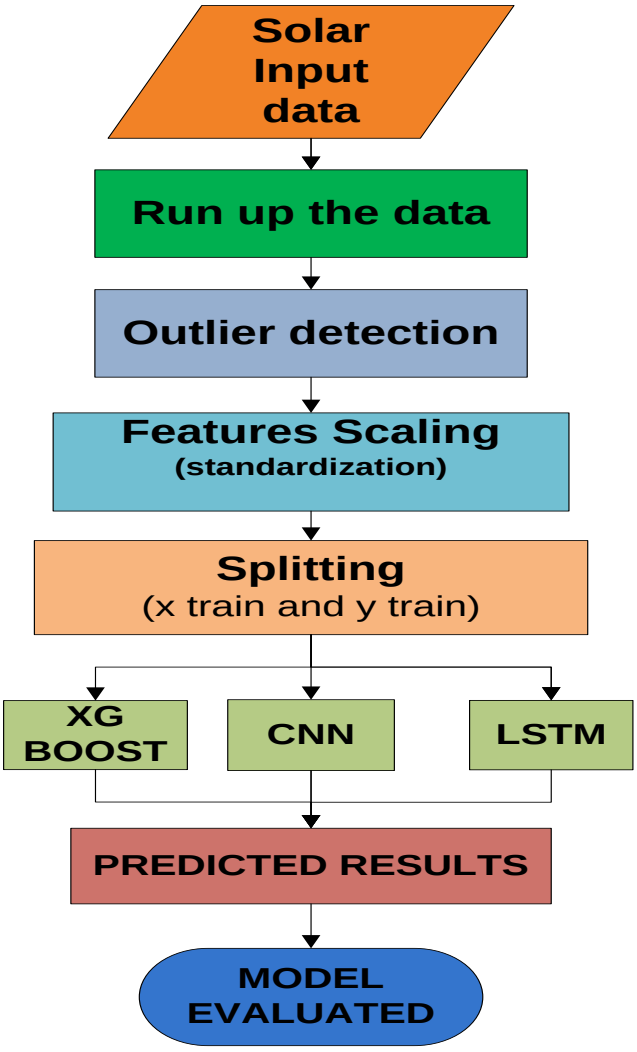


Figure 4.1: Flowchart for Processing of Proposed Models

The standardized data is then split into input features (X_train) and target outputs (Y_train), preparing it for model training. Three different machine learning models are employed for prediction: XG Boost, CNN and LSTM Each model is trained using the prepared data and generates predicted results. Finally, these predicted results are compared with the actual values to evaluate the model's performance. Evaluation metrics such as RMSE, MAE, and R^2 score are used to determine the accuracy and reliability of each model. This workflow ensures a comprehensive and comparative approach to selecting the most suitable model for solar forecasting.

4.2 Error Metrics for Solar Irradiance Forecasting

Table 4.1: Error Metrics of Solar Irradiance Forecasting

METHODS	RMSE	MSE	MAE	R − SQUARE
XGBOOST	0.5168	0.0267	0.0371	0.9723
CNN	0.56	0.3146	0.3188	0.9676
LSTM	0.0984	0.0969	0.0557	0.9989

Table 4.1 compares the performance of three machine learning models XG Boost, CNN, and LSTM using four evaluation metrics: RMSE, MSE, MAE, and R^2. Among the models, LSTM demonstrates the best performance, achieving the lowest RMSE (0.0984), MSE (0.0969), and MAE (0.0557), indicating minimal prediction errors. Additionally, it has the highest R^2 value (0.9989), meaning it explains nearly all the variance in the data. XG Boost performs moderately well, with an RMSE of 0.5168, an MSE of 0.0267, and an MAE of 0.0371, along with an R^2 value of 0.9723, showing reliable but slightly less accurate predictions than LSTM. On the other hand, CNN has the highest error values, with an RMSE of 0.56, an MSE of 0.3146, and an MAE of 0.3188, as well as the lowest R^2 (0.9676), indicating relatively weaker predictive performance. Overall, LSTM emerges as the most effective model for forecasting, followed by XG Boost, while CNN performs the least accurately.

4.3 Scatter graph for the proposed methodology

Fig. 4.2 given graph illustrates the performance comparison of three machine learning models XG Boost, CNN, and LSTM for forecasting Global Horizontal Irradiance (GHI) in Goa using scatter plots. Each plot represents the relationship between the actual GHI values (on the x-axis) and the predicted GHI values (on the y-axis). In the case of the XG Boost model, the scatter plot shows that the predicted values are moderately aligned with the diagonal reference line ($y = x$), indicating a fair level of accuracy. However, some deviation is observed, particularly in the mid-range GHI values, suggesting that XG Boost struggles to consistently capture the variation in solar irradiance.

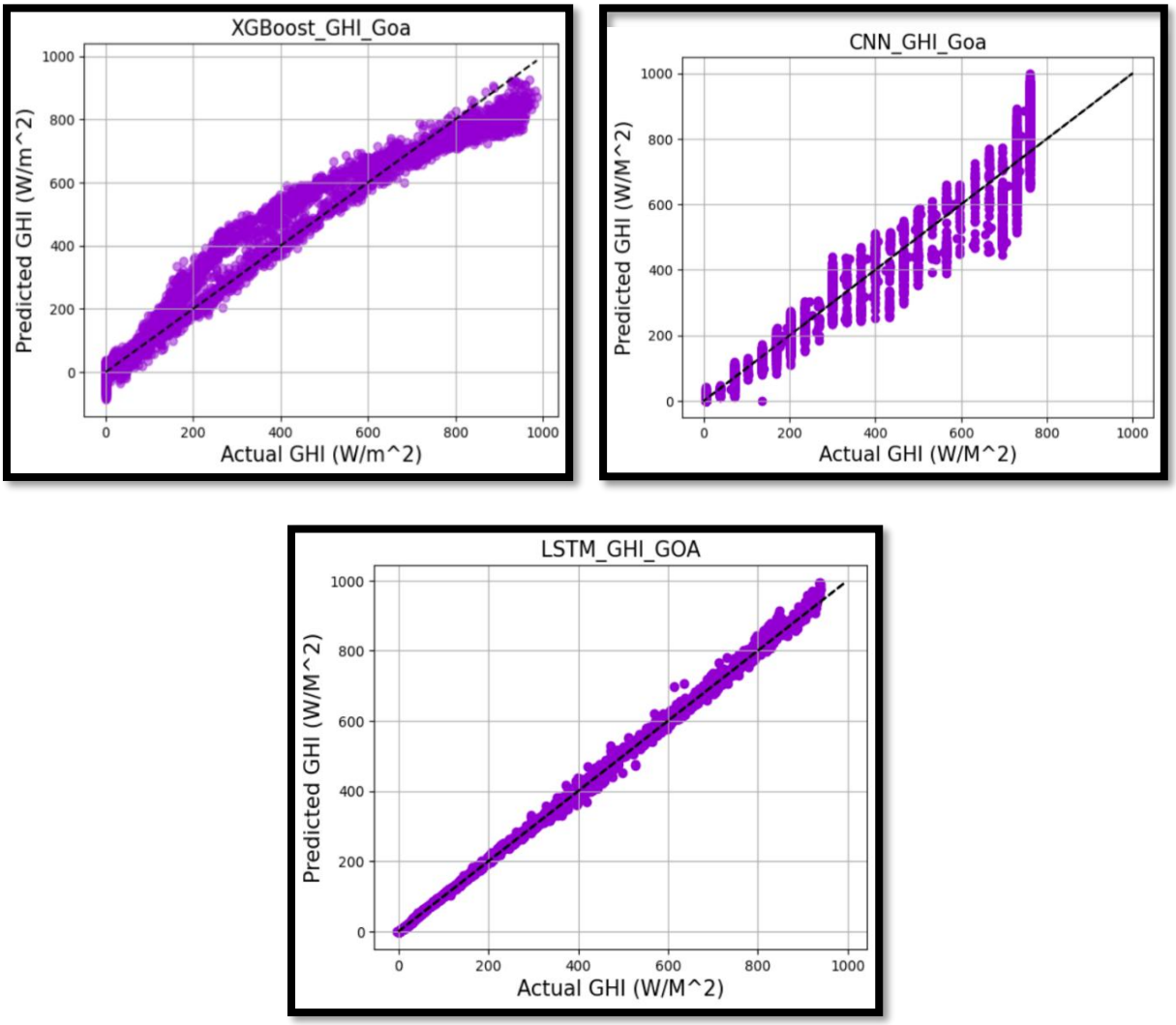


Figure 4.2: Scatter graph for Solar Irradiance Forecasting for (a)XG BOOST, (b) CNN (c) LSTM

Fig. 4.2 given graph illustrates the performance comparison of three machine learning models XG Boost, CNN, and LSTM for forecasting Global Horizontal Irradiance (GHI) in Goa using scatter plots. Each plot represents the relationship between the actual GHI values (on the x-axis) and the predicted GHI values (on the y-axis). In the case of the XG Boost model, the scatter plot shows that the predicted values are moderately aligned with the diagonal reference line (y = x), indicating a fair level of accuracy. However, some deviation is observed, particularly in the mid-range GHI values, suggesting that XG Boost struggles to consistently capture the variation in solar irradiance.

In contrast, the CNN model's scatter plot reveals a wider spread of data points around the diagonal line, especially at higher GHI values. This indicates a lower prediction accuracy and suggests that CNN may not be well-suited for capturing temporal patterns in solar data. Finally, the LSTM model demonstrates superior performance, as evidenced by the tightly clustered points along the diagonal. This high level of alignment reflects strong agreement between actual and predicted values. Due to its ability to model sequential dependencies in time-series data, LSTM delivers the most accurate and consistent results. Overall, the graph highlights that among the three models, LSTM performs the best for solar GHI forecasting in Goa, followed by XG Boost, while CNN shows the least accuracy. This suggests that LSTM is the most suitable model for time-series prediction in solar energy applications. The graph visually compares how well each model predicts solar GHI values for Goa. The closer the points are to the diagonal line, the more accurate the prediction. Among the three models:

1. LSTM performs best, showing high prediction accuracy.
2. XG Boost gives fair results with moderate errors.
3. CNN performs the least effectively, especially for high GHI values.

This visualization supports selecting **LSTM as the most reliable model** for solar forecasting in Goa.

CHAPTER 5

ENERGY MANAGEMENT IN ISOLATED MICROGRID

CHAPTER 5

ENERGY MANAGEMENT IN ISOLATED MICROGRID

5.1 Energy Management Through Energy Vector

With the increasing demand for reliable and sustainable energy solutions, isolated microgrids have emerged as a promising approach to provide power in remote or off-grid areas. These microgrids integrate renewable energy sources such as solar photovoltaic (PV) panels and wind turbines to harness clean energy, reducing dependence on conventional fossil fuels and minimizing environmental impact. However, the intermittent and variable nature of solar and wind energy presents significant challenges in ensuring a continuous and stable power supply, especially for critical loads that require uninterrupted operation.

To address these challenges, Battery Energy Storage Systems (BESS) play a crucial role in isolated microgrids by storing excess energy generated during peak production periods and supplying it during low-generation or high-demand periods. Effective energy management strategies are essential to optimize the coordination between renewable energy generation, battery storage, and load demand, thereby enhancing system reliability, efficiency, and lifespan of the battery.

This chapter focuses on the design and implementation of an energy management framework for an isolated microgrid that leverages solar and wind energy sources alongside battery storage to fulfil both general load and critical load requirements. It explores forecasting techniques for renewable energy generation, load demand analysis, and battery state-of-charge management to ensure uninterrupted power supply and maximize renewable energy utilization. The chapter also discusses optimization methods and control strategies that can improve the operational performance of the microgrid, making it a sustainable and resilient energy solution for isolated communities.

5.2 Load Demand forecasting

Load demand forecasting is the process of predicting future electricity consumption over a certain period based on historical data and influencing factors such as weather and economic conditions. It is essential for ensuring that power generation matches the

demand, which helps maintain grid stability, optimize resource use, reduce operational costs, and prevent power outages. Accurate forecasting is especially important in systems integrating renewable energy sources like solar and wind, as it aids in managing their variability and scheduling energy storage. By using various machine learning models, load demand forecasting enables efficient planning and operation of power systems, ultimately supporting reliable and cost-effective energy delivery.

5.2.1 Why load forecasting is necessary

Efficient Resource Planning: Accurate forecasting ensures that power generation matches the demand, preventing underproduction (leading to outages) or overproduction (leading to wasted resources).

Cost Reduction: It helps in optimizing fuel use and operational costs by scheduling generation units efficiently, reducing unnecessary starts and stops of power plants.

Grid Stability: Balancing supply and demand maintain grid reliability and prevents frequency fluctuations or blackouts.

Integration of Renewable Energy: Forecasting is crucial for managing the variability of renewable sources like solar and wind, enabling better scheduling and storage utilization.

Demand Response Management: Enables utilities to design demand-side management programs, shifting or reducing loads during peak periods.

5.2.2 Flowchart of Load Demand Forecasting and battery for energy management

Electric Load forecasting (ELF) and battery energy management are crucial for maintaining a consistent, efficient, and dependable power supply, especially in integrated RE systems [36]. Effective ELF enables better planning and optimization of energy generation, transmission, and storage, allowing supply and demand to be balanced while minimizing energy waste. Because RE sources such as solar and wind are intermittent, detailed forecasting enables grid managers to anticipate changes and make adjustments to ensure stability [37]. However, even with perfect forecasting, power outages can still occur owing to unforeseen variations in energy supply and demand. This is where battery energy management comes into play, storing surplus energy during peak generating periods and releasing it when demand exceeds supply levels. A well-optimized battery energy

management system improves grid reliability, reduces reliance on fossil fuel backup power, and facilitates the integration of RE [38]. Hence, it is vital to integrate ELF with battery energy management to maximize energy usage, increase grid resilience, and support the transition to a sustainable and smart energy infrastructure.

ELF of commercial facilities in southern Goa, by employing state-of-the-art techniques such as XG Boost [39], Convolutional Neural Networks (CNN) [40], and Long Short-Term Memory (LSTM) networks [41], is instrumental for energy management systems to be optimally designed. Load prediction, being precise and accurate, ensures effective energy utilization, particularly storage devices like batteries and backup systems. These predictions assist in making sure energy storage is effectively utilized according to actual consumption levels.

Commercial loads show high fluctuations owing to changes in business operations, weather, and time of day, which are more acute than those for residential loads. As such, the models therefore attempt to predict such fluctuations accurately and taking into consideration the random nature of energy demand in commercial applications.

Additionally, factors including variation in business hours, size of workforce, and utilization of equipment account for the variation in load profiles. Hence, shorter learning periods are desirable because longer periods may not equate to better forecast accuracy.

In this work, data from the southern region of Goa is collected, and it is proposed that 20% of the total load demand will be supplied by solar photovoltaic (PV) and wind generation. Additionally, the intermittent nature of RE is managed through a BEMS, which handles charging and discharging to maintain a stable energy supply.

Fig.5.1 represents a machine learning-based approach for evaluating battery capacity by predicting power demand and generation. It begins with the load demand being gathered and pre-processed in the run-up data stage. To ensure data quality, outliers' detection is performed, followed by feature scaling (standardization) to normalize the dataset

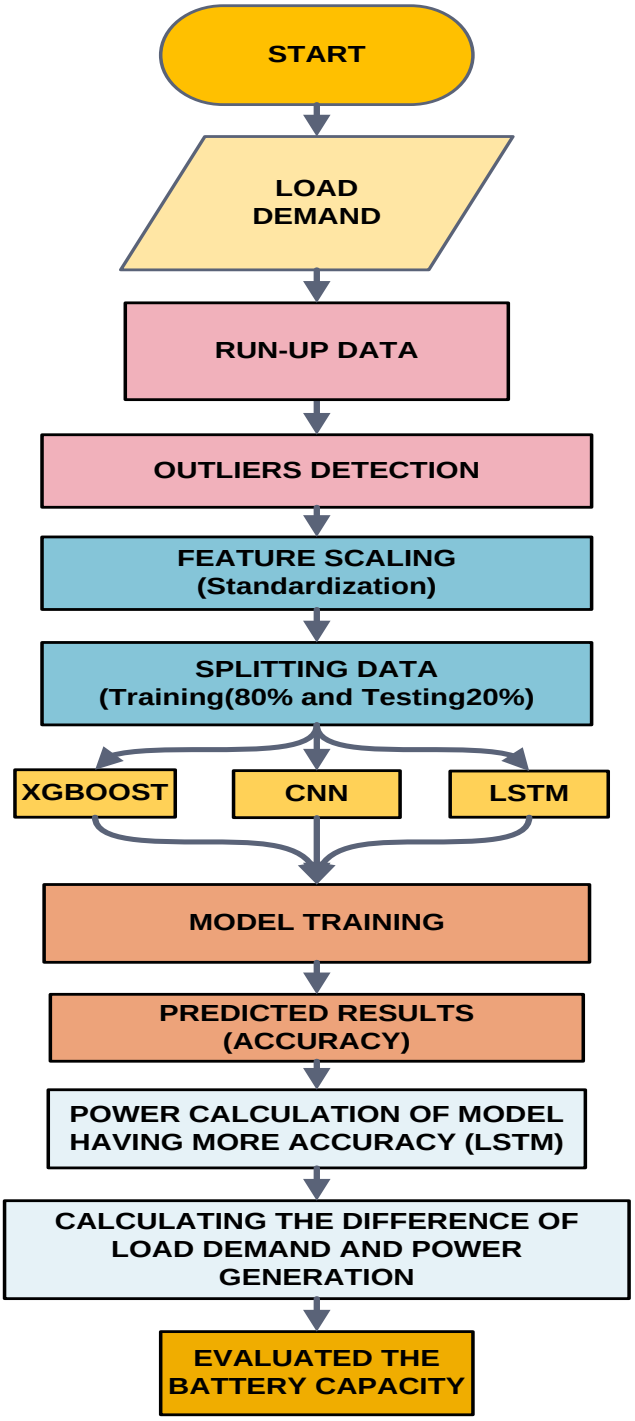


Figure 5.1: Flowchart of proposed Methodology

The data is then split into training 80% and testing 20% sets for model development. Three machine learning models XG Boost, CNN, and LSTM are considered for training, after which the predicted results (accuracy) are evaluated. The model with the highest accuracy,

typically LSTM, is used for power calculation. Subsequently, the difference between load demand and power generation is calculated, leading to the final step where the battery capacity is evaluated. This structured process ensures efficient energy demand forecasting and optimization of power storage systems.

5.2.3 Error Metrics for Load Demand Forecasting

Table 5.1: Performance Indices for Load Demand Forecasting for XG BOOST, CNN And LSTM

METHODS	MAE	RMSE	MSE	R- SQUARE
XG BOOST	0.4620	0.6355	0.0703	0.8954
CNN	0.4029	0.3526	0.0475	0.9151
LSTM	**0.2078**	**0.2742**	**0.0252**	**0.9242**

Table 5.1 presents a comparative analysis of three machine learning models XG Boost, CNN, and LSTM based on four performance metrics RMSE, MAE, MSE and R-SQUARE. Among the models, LSTM demonstrates the best performance with the lowest RMSE (0.2742), MAE (0.2078), and the highest R^2 value (0.9242), indicating that it provides the most accurate predictions. CNN follows closely with an RMSE of 0.3526 and an R^2 of 0.9151, showing competitive performance. In contrast, XG Boost has the highest RMSE (0.6355) and MAE (0.4620), signifying lower accuracy compared to the deep learning models. The MSE values further support this trend, where CNN (0.0475) and LSTM (0.0252) outperform XG Boost (0.0703). Overall, LSTM emerges as the most effective model for the given task, followed by CNN, while XG Boost exhibits relatively lower predictive accuracy

5.2.4 Scatter graph of Load Demand Forecasting

Fig 5.2 scatter plots presented above compare the actual load demand (in kW/hr) with the predicted load demand generated by three different machine learning models: XG Boost, CNN, and LSTM Each point on the graph represents a prediction instance, and the dotted black line indicates the ideal case where predicted values match the actual values perfectly

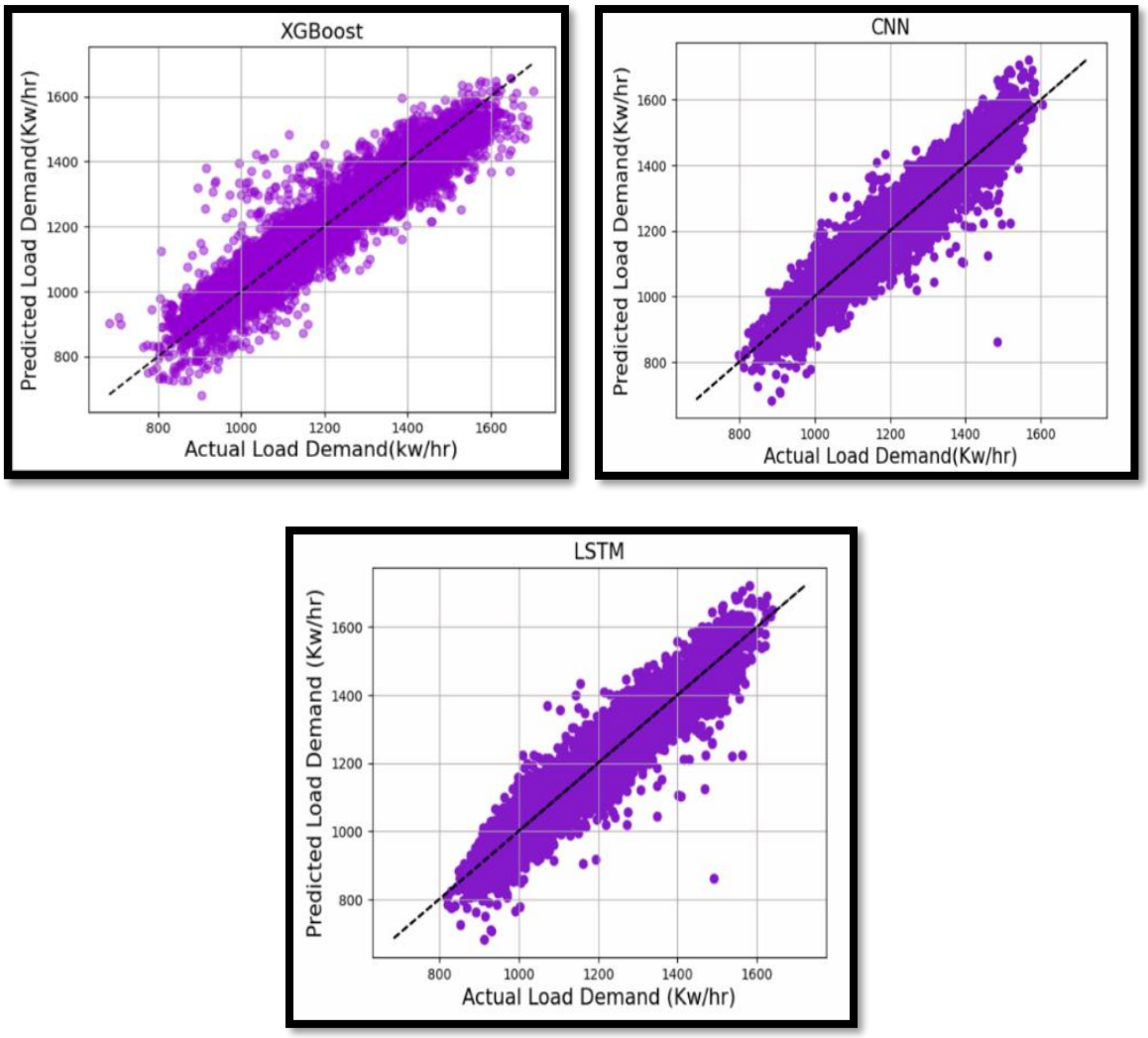


Figure 5.2: Scatter graph for Load Demand Forecasting for (a) XG BOOST (b) CNN (c) LSTM

Fig 5.2 scatter plots presented above compare the actual load demand (in kW/hr) with the predicted load demand generated by three different machine learning models: XG Boost, CNN, and LSTM Each point on the graph represents a prediction instance, and the dotted black line indicates the ideal case where predicted values match the actual values perfectly (i.e., a 1:1 line).Fig 5.2 (a) the predicted values are relatively close to the actual values, forming a tight cluster around the ideal line, though with slightly more spread at higher and lower demand values. This indicates that XG Boost performs fairly well in capturing the load demand trend but may introduce some prediction error in extreme values. Fig 5.2 (b) demonstrates a more compact clustering of points along the diagonal, especially in the mid-range of load demand. This indicates better prediction consistency and less deviation compared to XG Boost. However, some outliers are still present, mainly at the edges.

Fig 5.2(c) shows the highest concentration of points along the diagonal, suggesting the most accurate and stable predictions among the three models. The predictions by LSTM are very close to the actual values across all ranges, with minimal scattering and fewer outliers. While all three models exhibit good prediction performance, LSTM outperforms CNN and XG Boost in terms of accuracy and robustness in forecasting load demand.

5.3 Hybrid Renewable Energy System Integrating Solar, Wind and Battery Storage

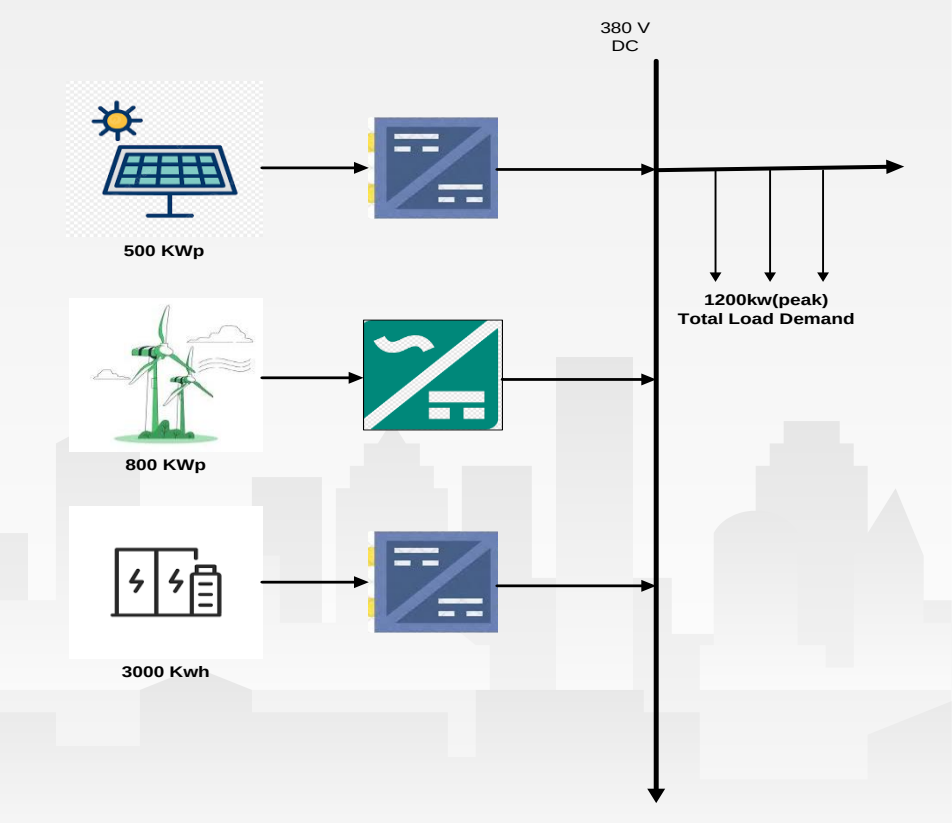


Figure 5.3: Represents the hybrid renewable energy system integrating solar, wind and battery storage.

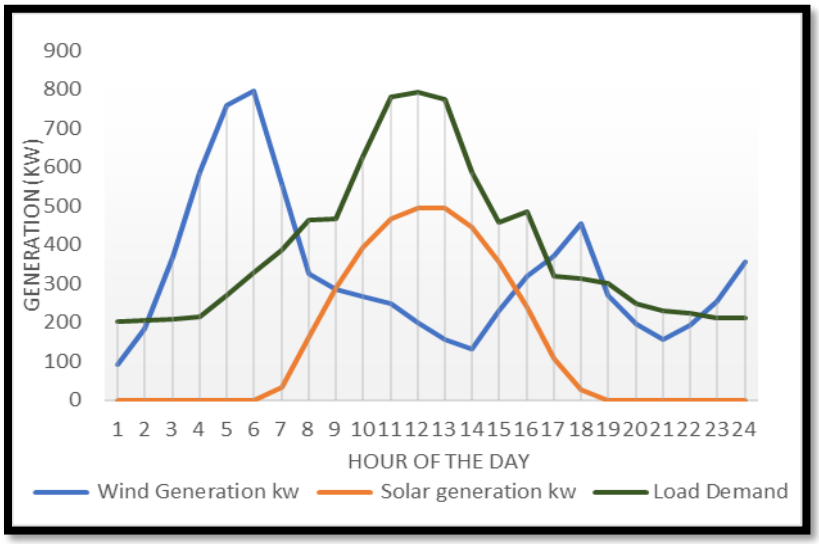
Fig 5.3 illustrates a hybrid renewable energy system designed to meet a peak total load demand of 1200 kW at 380 V DC. It integrates three primary energy sources:

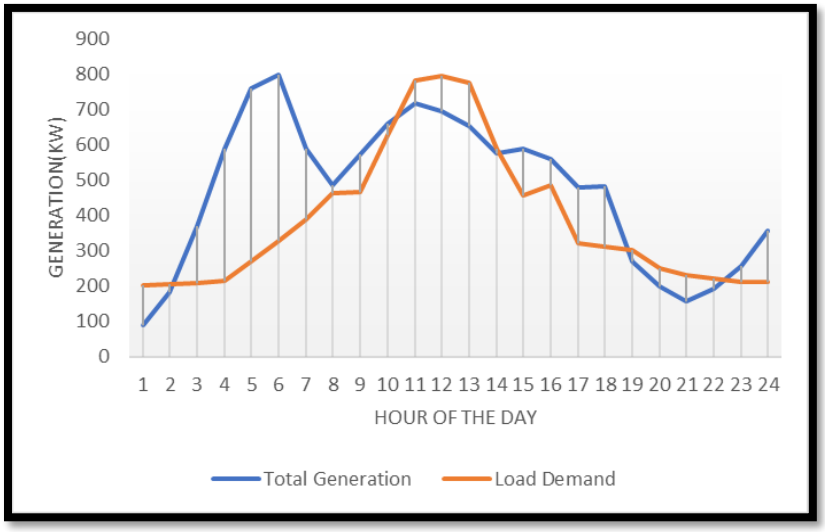
- **Solar Power:** A PV system rated at 500 kWp supplies power through a DC-DC converter, feeding directly into the DC bus.
- **Wind Power:** A wind energy system rated at 800 kWp supplies power via an AC-DC converter, indicating that the wind turbines generate AC which is converted to DC for integration into the system.

- **Battery Energy Storage System (BESS):** A battery bank with a storage capacity of 3000 kWh provides backup and load balancing. It connects to the DC grid through a bidirectional DC-DC converter, allowing both charging and discharging operations.

All three sources feed into a common 380 V DC bus, which then supplies power to the total load, peaking at 1200 kW. This setup ensures a stable power supply by utilizing renewable sources and storage, contributing to energy reliability and sustainability in commercial or industrial applications.

5.3.1 Graphical representation of solar, wind generation with load demand profile for 24 hrs.





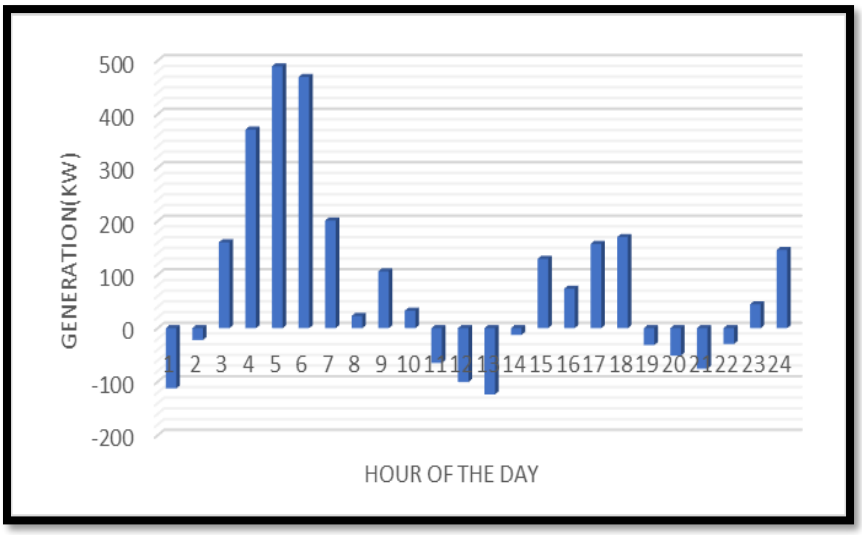


Figure 5.4 Graphical presentation of (a)solar wind generation with load demand profile for 24 hours(b) total generation and load demand for 24 hr period (c) deficit power

Fig. 5.4 (a) shows wind generation, solar generation, and load demand over 24 hours. Wind energy peaks in the early morning and evening, while solar energy dominates midday. Load demand rises after 8 AM, peaks around noon, and again in the evening. Wind and solar complement each other, but mismatches between generation and demand highlight the need for energy storage or backup sources. This analysis is essential for optimizing renewable energy use and ensuring grid stability. Fig.5.4 (b) compares total renewable energy generation (blue line) with load demand (orange line) over 24 hours. Generation peaks in the early morning and midday, aligning well with demand during solar hours but falling short during certain periods, especially at night. There are instances where generation exceeds demand, indicating potential for energy storage. Fig.5.4 (c) shows the difference between total generation and load demand over 24 hours. There is a surplus of energy in the early morning (3-7 AM), a balance around midday (9 AM-1 PM), and a shortfall in the afternoon (2-4 PM). Evening hours show fluctuations between surplus and shortage. This highlights the need for energy storage and backup sources to maintain a stable power supply.

5.4 State of Charge of Battery Energy Storage for Three Initial SOC Condition

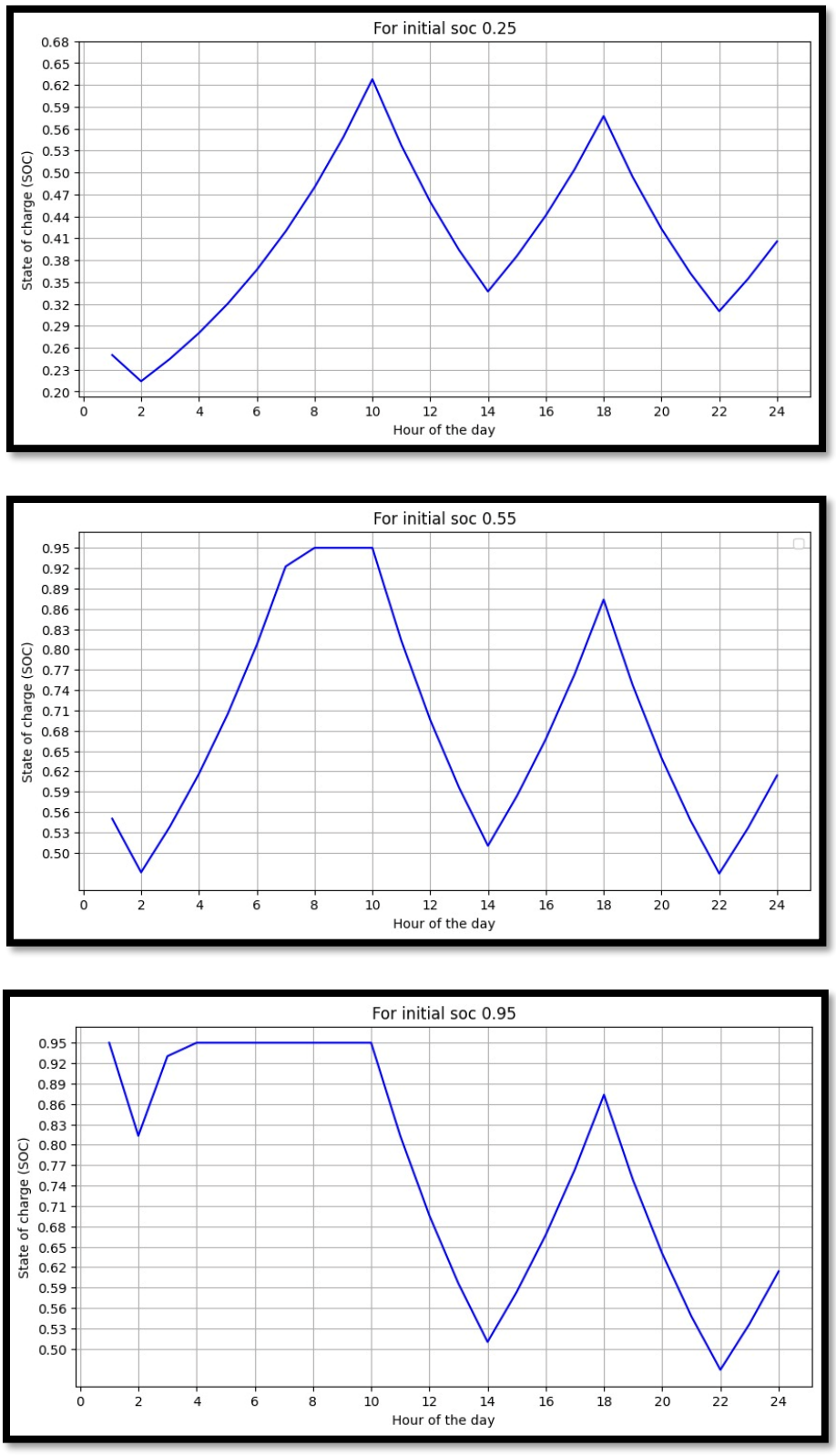

For initial soc 0.25
State of charge (SOC)
Hour of the day
For initial soc 0.55
State of charge (SOC)
Hour of the day
For initial soc 0.95
State of charge (SOC)
Hour of the day

Figure 5.5: SoC graph of BES for three initial SOC conditions (a) 0.25, (b)0.5 and (c) 0.95

5.4.1 Case 1: Soc for 0.25

Fig 5.5(a) shows a 24-hour profile of the SoC of a battery energy storage system, starting with an initial SoC of 0.25 (25%). The graph helps analyze how the battery charges and discharges throughout the day in response to load demand and power generation from renewable sources**.**

Axes Explanation:

- **X-axis:** Represents the Hour of the Day (from 0 to 24).
- **Y-axis:** Shows SoC as a fraction (ranging approximately from 0.2 to 0.7).

Observations:

Initial Dip (Hours 0–2): The SoC slightly drops after midnight, indicating energy discharge—likely due to demand exceeding generation at night.
Morning Rise (Hours 2–10): The SoC steadily increases, peaking around hour 10. This suggests solar and/or wind generation is charging the battery during the morning hours.
Midday Drop (Hours 10–14): A sharp decrease in SoC implies a significant discharge, likely due to a rise in load demand or reduced generation.
Afternoon Recovery (Hours 14–18): The battery charges again, possibly due to solar power generation resuming or decreasing load demand.
Evening Discharge (Hours 18–22): A second major drop in SoC occurs, consistent with peak evening load periods when stored energy is used to meet demand.
Late-Night Stabilization (Hours 22–24): A slight recovery in SoC hints at reduced consumption or minor charging during night hours.

5.4.2 Case 2: Soc for 0.55

Fig 5.5(b) illustrates the SoC of a BESS over a 24-hour period, with an **initial SoC of 0.55 (55%)**. It highlights how the battery charges and discharges in response to energy generation and load demand throughout the day.

Axes Details:

- **X-axis**: Hour of the Day (0–24)
- **Y-axis**: SoC ranging from 0.5 to nearly 1.0

Observations:

Early Hours (0–2): The SoC decreases slightly, indicating battery discharge during low generation hours (likely nighttime load support).

Morning Charge (2–8): A steady increase in SoC occurs, reaching **0.95**, suggesting substantial solar or wind input. This is possibly the time of high renewable energy availability (morning sun or wind).

Saturation Point (8–10): The SoC remains constant at **0.95**, likely because the battery is near full capacity and cannot store more energy, even though generation continues.

Midday to Early Afternoon Discharge (10–14): A steep decline in SoC indicates heavy battery usage to meet peak daytime load or reduced renewable input.

Afternoon Recharge (14–18): Battery charges again, showing another generation phase—possibly afternoon sun or wind supply.

Evening Discharge (18–22): SoC again drops sharply, indicating peak evening demand is being supported by the battery.

Late-Night Stabilization (22–24): A minor recovery in SoC shows that load decreases, or minimal generation resumes late at night.

5.4.3 Case 3: soc for 0.95

Fig 5.4(c) displays the SoC of a BESS over a 24-hour period, starting with an initial SoC of 0.95 (95%). It represents how the battery is managed throughout the day to balance energy generation and consumption in an isolated or hybrid system.

Axes Information:

X-axis: Hour of the Day (0–24)

Y-axis: SoC from 0.5 to nearly 1.0

Observations:

Initial High SoC (Hour 0–1): The battery starts nearly full at 0.95 and experiences a slight discharge around hour 2, indicating usage during nighttime demand.

Quick Recharge (Hour 2–4): SoC rises back to 0.95, showing that some generation (e.g., early wind or minimal solar) restores the battery to full capacity.

Plateau Phase (Hour 4–10): The battery remains fully charged at 0.95, indicating either minimal load or continuous generation with limited need for discharging.

Gradual Discharge (Hour 10–14): A sharp decrease in SoC suggests high load demand

or reduced energy input. The battery provides energy support during peak load hours.

Afternoon Recharge (Hour 14–18): The SoC increases again, possibly due to solar or wind generation in the afternoon, partially recharging the battery.

Evening Discharge (Hour 18–22): Another steep fall in SoC is observed, indicating the battery is meeting evening peak load demands when generation is low or absent.

Late-Night Stabilization (Hour 22–24): The SoC rises slightly, indicating a drop in demand or a small contribution from renewable sources.

A higher initial SoC enhances energy availability and system reliability, enabling better load management and smoother SoC transitions. However, all cases reflect the importance of renewable generation timing in maintaining battery efficiency and ensuring uninterrupted power supply.

.

Chapter 6

CONCLUSION AND FUTURE SCOPE

CHAPTER 6

CONCLUSION AND FUTURE SCOPE

This research presents an integrated and in-depth analysis of wind speed, solar irradiance and load demand forecasting using a combination of Machine Learning (ML) and Deep Learning (DL) techniques. The primary objective is to ensure accurate renewable energy forecasting and efficient energy management in a grid-isolated energy system, with a special focus on Battery Energy Storage Systems (BESS) for handling intermittent power supply and demand fluctuations. The entire study is divided into three core segments: wind speed forecasting, renewable energy forecasting (solar and wind), and load demand prediction integrated with energy management strategies.

6.1 Conclusion

Part I: Wind Speed Forecasting Using Machine Learning Approaches

The first part of the study evaluates the performance of several traditional yet powerful ML algorithms including Support Vector Regression (SVR), Random Forest (RF), Decision Tree (DT), Multiple Linear Regression (MLR), and XG Boost for wind speed forecasting. A major emphasis is placed on the role of data preprocessing, which is shown to have a substantial impact on improving the accuracy and robustness of prediction models. The study employs three distinct case studies to explore various conditions and configurations.

In Case I, raw data without preprocessing is used to test the models. While SVR still achieves the highest R^2 value, the graphical results indicate inconsistency in aligning predicted values with actual ones, suggesting the need for better data handling techniques. Moving into Case II, after thorough data preprocessing including normalization, removal of outliers, and imputation of missing values the prediction accuracy of all models improves noticeably. SVR again stands out, with its performance metrics significantly enhanced. Post-preprocessing values for SVR show an RMSE of 0.3882, MSE of 0.1507, MAE of 0.2724, and an R^2 of 0.771, indicating improved model fitting and reduced error margins.

Case III introduces seasonal analysis, wherein forecasting models are tested separately for summer, rainy, and winter seasons. The results confirm that SVR continues to deliver the best performance across all seasons, followed by RF. The consistency of SVR's superior results across both annual and seasonal datasets suggests that it is a robust choice for wind energy forecasting in varying weather conditions. For instance, in summer, SVR achieves an RMSE of 0.4747, MSE of 0.2434, MAE of 0.3278, and an R^2 of 0.7711. Similar metrics are recorded for the rainy and winter seasons. This part of the study not only highlights the impact of preprocessing and algorithm choice but also illustrates how well-optimized ML models can adapt to different temporal patterns in wind data.

Part II: Renewable Energy Forecasting Using Deep Learning Models

The second phase of the research transitions to the application of advanced deep learning models, specifically Long Short-Term Memory (LSTM) networks, Convolutional Neural Networks (CNN), and XG Boost, for forecasting solar irradiance and wind speed. These models are chosen for their ability to capture complex temporal and spatial patterns, which are crucial for renewable energy systems due to their inherently fluctuating nature.

Among the three, LSTM emerges as the most accurate and reliable model, offering minimal prediction error and maximum correlation between predicted and actual values. It achieves a remarkable R^2 of 0.9989 for solar irradiance and 0.9684 for wind speed, underscoring its effectiveness in modelling time-dependent sequences. The LSTM model also demonstrates the lowest RMSE, MSE, and MAE values, confirming its superior predictive capacity over other DL models.

In contrast, CNN, while showing competitive performance in solar irradiance prediction, records a relatively higher MAE, indicating less consistency in its outputs. XG Boost, although known for its speed and efficiency in various forecasting tasks, underperforms in this context with higher error margins and lower R^2 values, making it less suitable for precise renewable forecasting in critical energy systems. The results affirm that LSTM is highly suitable for forecasting both solar and wind components of renewable energy, which plays a crucial role in optimizing the planning and operation of energy systems.

Part III: Load Demand Forecasting and Energy Management with BESS

In the final segment, the study focuses on forecasting electrical load demand and designing an efficient energy management system using the results from the renewable energy forecasts. Once again, LSTM demonstrates exceptional predictive performance, outperforming CNN and XG Boost with the lowest error metrics and the highest R^2 values. This confirms LSTM's versatility and robustness in modelling not only renewable energy generation but also variable load demand profiles.

From the forecasting analysis, the peak load demand is identified as 1,180 kW. To meet this demand, a hybrid renewable energy system is proposed, consisting of an 800-kW wind turbine and a 500-kW solar PV system, yielding a total generation capacity of 1,300 kW. This additional 120 kW margin is intentionally designed to accommodate the intermittent and variable nature of renewable resources.

Despite the surplus capacity, the system analysis reveals that power deficits may still arise during periods of reduced generation or sudden load spikes. To ensure continuous power availability and system stability, a 125 kW Battery Energy Storage System (BESS) is integrated. The State of Charge (SOC) of the battery is evaluated under three different initial charge conditions to assess its ability to respond to load variations. The results demonstrate that the BESS effectively compensates for the power shortfalls, enhancing the reliability and resilience of the isolated energy system.

This segment validates the significance of combining accurate forecasting models with strategic resource planning and storage solutions to create self-sustaining, grid-independent power systems.

6.2 Future Scope

This comprehensive study highlights the critical role of advanced data analytics in enhancing the reliability and efficiency of renewable energy systems, particularly in grid-isolated environments. It demonstrates that:

- ML techniques like SVR and RF are well-suited for short-term and seasonal forecasting when properly optimized with data preprocessing.

- DL models, especially LSTM, offer outstanding performance for both resource and load forecasting, owing to their ability to model long-term temporal dependencies.
- The integration of hybrid renewable energy sources with BESS, supported by precise forecasting, ensures stable power supply even in isolated or off-grid scenarios.

The insights derived from this study provide a solid foundation for real-world implementation of renewable energy forecasting and management systems. Moreover, the findings encourage the exploration of hybrid and ensemble forecasting models, which could leverage the strengths of multiple algorithms to further improve accuracy. Future research may also consider real-time control strategies, dynamic load adjustment, and demand-side management integrated with smart grid technologies to further enhance the autonomy and resilience of grid-isolated systems.

REFERENCES

1. Y. Zhang, N. Rahbari-Asr, J. Duan, and M.-Y. Chow, “Day-ahead smart grid cooperative distributed energy scheduling with renewable and storage integration,” IEEE Transactions on Sustainable Energy, vol. 7, no. 4, pp. 1739–1748, 2016.
2. D. Markovics and M. J. Mayer, “Comparison of machine learning methods for photovoltaic power forecasting based on numerical weather prediction,” Renewable and Sustainable Energy Reviews, vol. 161, p. 112364, 2022.
3. Y. Zhang, H. Sun, and Y. Guo, “Wind power prediction based on pso-svr and grey combination model,” IEEE Access, vol. 7, pp. 136254–136267, 2019.
4. M. Fatnani, D. Naware, and A. Mitra, “Design of solar pv based ev charging station with optimized battery energy storage system,” in 2020 IEEE First International Conference on Smart Technologies for Power, Energy and Control (STPEC), pp. 1–5, 2020.
5. B. Tan, X. Ma, Q. Shi, M. Guo, H. Zhao, and X. Shen, “Ultra-shortterm wind power forecasting based on improved lstm,” in 2021 6th International Conference on Power and Renewable Energy (ICPRE), pp. 1029–1033, 2021.
6. M. B. Ozkan and P. Karagoz, “A novel wind power forecast model: Statistical hybrid wind power forecast technique (shwip),” IEEE Transactions on Industrial Informatics, vol. 11, no. 2, pp. 375–387, 2015.
7. L. Visser, T. AlSkaif, and W. van Sark, “Operational day-ahead solar power forecasting for aggregated pv systems with a varying spatial distribution,” Renewable Energy, vol. 183, pp. 267–282, 2022.
8. Brahimi, T. Using Artificial Intelligence to Predict Wind speed for Energy Application in Saudi Arabia. Energies 2019, 12, 4669.
9. Nomanbhay, R. Hussein, M.Y. Ong, Sustainability of biodiesel production in Malaysia by production of bio-oil from crude glycerol using microwave pyrolysis: a review, Green Chem. Lett. Rev. 11 (2) (2018) 135–157.
10. Dora, B.K., Bhat, S., Halder, S. and Srivastava, I., 2024. A solution to multi Objective Stochastic Optimal Power Flow problem using mutualism and elite strategy based Pelican Optimization Algorithm. *Applied Soft Computing*, *158*, p.111548
11. M.N.I. Maruf, Open model-based analysis of a 100 % renewable and sector-coupled energy system–the case of Germany in 2050, Appl. Energy 288 (2021), 116618.
12. N.E.F. Bloomberg, New Energy Outlook 2021-Executive Summary. 2021. Available from: 〈https://about.bnef.com/new-energy-outlook〉
13. M. Fatnani, D. Naware and A. Mitra, "Design of Solar PV Based EV Charging Station with Optimized Battery Energy Storage System," 2020 IEEE First International Conference on Smart Technologies for Power, Energy and Control (STPEC).

14. R. S. Waghmare, S. Bhat and B. K. Dora, "Contingency Analysis of IEEE 14 Bus System with Integration of DFIG Wind Farm," *2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET)*, Hyderabad, India, 2024, pp. 1-5.

15. Kariniotakis, G.N.; Stavrakakis, G.S.; Nogaret, E.F. Wind power forecasting using advanced neural networks models. IEEE Trans. Energy Convers. 1996, 11, 762–767.

16. Wang, Y.; Wu, L. On practical challenges of decomposition-based hybrid forecasting algorithms for wind speed and solar irradiation. Energy 2016, 112, 208–220.

17. Z.-X. Wang, Z.-W. Wang, Q. Li, Forecasting the industrial solar energy consumption using a novel seasonal GM(1,1) model with dynamic seasonal adjustment factors, Energy 200 (2020), 117460.

18. Dora, B.K., Bhat, S., Halder, S. and Srivastava, I., 2024. A solution to multi Objective Stochastic Optimal Power Flow problem using mutualism and elite strategy based Pelican Optimization Algorithm. *Applied Soft Computing*, *158*, p.111548.

19. McCollum, D.L., Zhou, W., Bertram, C., De Boer, H.S., Bosetti, V., Busch, S., Després, J., Drouet, L., Emmerling, J., Fay, M. and Fricko, O., 2018. Energy investment needs for fulfilling the Paris Agreement and achieving the Sustainable Development Goals. *Nature Energy*, *3*(7), pp.589-599.

20. Dora, B.K.; Bhat, S.; Mitra, A.; Ernst, D.; Halinka, A.; Zychma, D.; Sowa, P. The Global Electricity Grid: A Comprehensive Review. *Energies* 2025, *18*, 1152.

21. Yadav, N.K., Jangilwar, P.K. and Das, S., 2024, July. Forecasting of Wind Power Using Modified Exponential Gaussian Process Regression. In *2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET)* (pp. 1-6). IEEE.

22. Payal, J., Das, S. and Bhukya, J., 2024, July. Impact of Hyperparameter Tunning in Hourly Solar Irradiance Forecasting by Using Machine Learning Techniques. In *2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET)* (pp. 1-6). IEEE.

23. Liao, W., Wang, S., Yang, D., Yang, Z., Fang, J., Rehtanz, C. and Porté-Agel, F., 2025. TimeGPT in load forecasting: A large time series model perspective. *Applied Energy*, *379*, p.124973.

24. Hnin, S.W., Karnjana, J., Kohda, Y. and Jeenanunta, C., 2024. A hybrid K-means and KNN approach for enhanced short-term load forecasting incorporating holiday effects. *Energy Reports*, *12*, pp.5942-5959.

25. Hua, P., Wang, H., Xie, Z. and Lahdelma, R., 2024. District heating load patterns and short-term forecasting for buildings and city level. *Energy*, *289*, p.129866.

26. X. Liu, B. Moreno, A.S. García, A grey neural network and input-output combined forecasting model. Primary energy consumption forecasts in Spanish economic sectors, Energy 115 (2016) 1042–1054

27. T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2016, pp. 785–794.

28. H. Lu, et al., Short-term prediction of building energy consumption employing an improved extreme gradient boosting model: a case study of an intake tower, Energy 203 (2020), 117756.

29. S. Touzani, J. Granderson, S. Fernandes, Gradient boosting machine for modeling the energy consumption of commercial buildings, Energy Build. 158 (2018) 1533–1543.

30. B. Tan, X. Ma, Q. Shi, M. Guo, H. Zhao, and X. Shen, "Ultra-shortterm wind power forecasting based on improved lstm," in 2021 6th International Conference on Power and Renewable Energy (ICPRE), pp. 1029–1033, 2021.

31. T. Chen, C. Guestrin, XGBoost: a scalable tree boosting system, in: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 2016, pp. 785–794.

32. M. B. Ozkan and P. Karagoz, "A novel wind power forecast model: Statistical hybrid wind power forecast technique (shwip)," IEEE Transactions on Industrial Informatics, vol. 11, no. 2, pp. 375–387, 2015.

33. Salman, D., Direkoglu, C., Kusaf, M. *et al.* Hybrid deep learning models for time series forecasting of solar power. *Neural Comput & Applic* **36**, 9095–9112 (2024).

34. J. Payal, S. Das, and J. Bhukya, "Impact of hyperparameter tunning in hourly solar irradiance forecasting by using machine learning techniques," in 2024 IEEE 4th International Conference on Sustainable Energy and Future Electric Transportation (SEFET), pp. 1–6, 2024.

Kariniotakis, G.N.; Stavrakakis, G.S.; Nogaret, E.F. Wind power forecasting using advanced neural networks models. IEEE Trans. Energy Convers. 1996, 11, 762–767

35. Naware, D., Badigenchala, R., Mitra, A. and Das, D., 2022. Impact of demand response on battery energy storage degradation using gbest-guided artificial bee colony algorithm with forecasted solar insolation. *Journal of Energy Storage*, *52*, p.104915.

36. Ma, P., Cui, S., Chen, M., Zhou, S. and Wang, K., 2023. Review of family-level short-term load forecasting and its application in household energy management system. *Energies*, *16*(15), p.5809.

37. Yadav, N.K. and Das, S., 2025. Multi-Objective Optimization for Distributed Generator and Shunt Capacitor Placement Considering Voltage-Dependent Nonlinear Load Models. *Swarm and Evolutionary Computation*, *92*, p.101782.

38. Choi, M.E., Kim, S.W. and Seo, S.W., 2011. Energy management optimization in a battery/supercapacitor hybrid energy storage system. *IEEE Transactions on Smart grid*, *3*(1), pp.463-472.

39. Wang, Y., Sun, S., Chen, X., Zeng, X., Kong, Y., Chen, J., Guo, Y. and Wang, T., 2021. Short-term load forecasting of industrial customers based on SVMD and XGBoost. *International Journal of Electrical Power & Energy Systems*, *129*, p.106830.

40. Rafi, S.H., Deeba, S.R. and Hossain, E., 2021. A short-term load forecasting method using integrated CNN and LSTM network. *IEEE access*, *9*, pp.32436-32448.

41. Chen, T. and Guestrin, C., 2016, August. Xgboost: A scalable tree boosting system. In *Proceedings of the 22nd acm sigkdd international conference on knowledge discovery and data mining* (pp. 785-794).